

TARTU ÜLIKOOL
Majandusteaduskond

Märt Paal

**PANKROTI PROGNOOSIMISE MUDELITE
KOOSTAMINE EUROOPA
AUTOKAUBANDUSE ETTEVÕTETE NÄITEL**

Magistritöö sotsiaalteaduse magistrikraadi taotlemiseks majandusteaduses

Juhendaja: teadur Oliver Lukason

Tartu 2016

Soovitan suunata kaitsmisele
Oliver Lukason

Kaitsmisele lubatud “ 2016. a

Olen koostanud töö iseseisvalt. Kõik töö koostamisel kasutatud teiste autorite tööd, põhimõttelised seisukohad, kirjandusallikatest ja mujalt pärinevad andmed on viidatud.

.....
Märt Paal

SISUKORD

Sissejuhatus.....	3
1.ETTEVÕTTE PANKROTI PROGNOOSIMISE TEOREETILISED ALUSED	6
1.1.Ettevõtte pankrot ja selle prognoosimise meetodid	6
1.2.Kaubandussektori pankroti prognoosimisel kasutatavad muutujad.....	21
2.EUROOPA AUTOKAUBANDUSETTEVÕTETE PANKROTI PROGNOOSIMISE MUDELITE KOOSTAMINE JA VÕRDLUS.....	33
2.1.Mudelite koostamiseks kasutatud andmed ja metoodika.....	33
2.2.Euroopa autokaubanduse ettevõtete pankrotistumise prognoosimine logistilise regressiooni meetodil	41
2.3.Euroopa autokaubanduse ettevõtete pankrotistumiste modelleerimine masinõppimise meetodeid rakendades	46
2.4.Erinevatel modelleerimise meetoditel saadud tulemuste võrdlus.....	53
Kokkuvõte.....	58
Viidatud allikad.....	62
Summary	68

SISSEJUHATUS

Ettevõtete pankrotistumise prognoosimise uurimine praegusel kujul sai alguse Altmani 1968. aastal ilmunud artikliga ning sealt tuttava Z-skooriga. Enamasti on koostatud mudelid keskendunud tööstusettevõtetele või äärmisel juhul on tegu erinevatele sektorite segamudelitega. Kaubandussektori pankroti prognoosimise mudeleid on koostatud üksikuid ning nende hulgast võib välja tuua järgmised: Bhargava ja teised (1998), McGurr ja Devaney (1998), Shirata (1998), Lukason (2006), Lee ja Choi (2013) ning käesoleva magistritöö autor enda bakalaureusetöös, milles koostas Eesti kaubanduse ettevõtete pankrotistumise prognoosimise mudeli logistilist regressiooni kasutades. Seega võib väita, et kaubandussektori ettevõtete pankrotistumise prognoosimist on varasemas teaduskirjanduses võrreldes tööstussektoriga (enamus ettevõtete pankrotistumise mudeleid on koostatud tööstussektori ettevõtete põhiselt) vähe uuritud. Autokaubanduse sektor on hästi piiritletav võrreldes teiste kaubanduse sektoritega ning on oma käibemahult oluliselt volatiilsem, kui kaubandussektor tervikuna, mis tähendab, et selle on ebaõnnestumisest rohkem ohustatud, kui keskmine kaubandussektori ettevõtte. Eelnevast lähtuvalt on autokaubanduse ettevõtete pankrotistumise uurimine aktuaalne.

Antud magistritöö keskendub Euroopa autokaubanduse sektori spetsiifiliste pankrotistumis-mudelite koostamisele. Seda sektorit pole varem ettevõtete pankrotistumise prognoosimise seisukohast uurinud ning see andis võimaluse koostada unikaalne magistritöö. Uuritava sektori valimise kriteeriumiks on piiritletavus. See võimaldab koostatavatest mudelitest leida sellele sektorile spetsiifilisi näitajaid, mis kõige paremini antud sektori ettevõtete ebaedu kirjeldavad. Sektor peaks olema ka

piisavalt suure müügikäibega — autokaubandus moodustades 2014. aasta Euroopa Liidu kaubandusettevõtete käibest 11% (Eurostat, autori arvutused).

Magistritöö eesmärgiks on koostada erinevatel meetoditel Euroopa autokaubanduse ettevõtete põhiseid pankrotistumise prognoosimise mudeleid ja leida nendest kontrollvalimi alusel kõige suurema õigestiklassifitseerimise määraga meetod. Eesmärgiga kaasnevalt on võimalik välja tuua sektori kõige olulisemad finantsnäitajad, siinkohal finantssuhtarvud autokaubanduse sektori pankrotistumise kirjeldamiseks. Seatud peaesmärgini jõudmiseks on antud magistritöö autor püstitanud järgmised uurimisülesanded:

1. käsitleda ettevõtete ebaõnnestumise ja pankrotistumise mõistet varasemate uurimuste näitel;
2. tutvustada varasemalt kasutatud pankrotistumise prognoosimise meetodeid;
3. tuua ülevaade ettevõtete pankrotistumise prognoosimise mudelites varem kasutatud finantssuhtarvudest fookusega kaubandussektorile;
4. anda lühike ülevaade autokaubanduse sektorist;
5. koostada autokaubanduse pankrotistumise prognoosimise mudelid erinevaid metoodikaid kasutades;
6. analüüsida ning võrrelda erinevate meetodite abil saadud tulemusi.

Magistritöös kasutab autor klassikalise statistilise meetodina logistilist regressiooni. Masinõppimise meetodeid on varasemalt ettevõtete pankrotistumise prognoosimisel vähe kasutatud. Antud töös võtab autor kasutusele k-lähima naabri meetodi ning juhusliku metsa meetodi.

Autokaubanduse sektori ülevaate andmisel teeb autor Euroopa Liidu autokaubandus-sektorist ülevaate, mis koosneb haru 2010...2014 perioodi ettevõtete arvu ning müügikäibe dünaamika analüüsist. Mudelite koostamisel kasutatav andmebaas pärineb Bureau van Dijk'i Amadeusi andmebaasist. Andmebaasi koostamiseks kasutab autor mootor sõidukite ja mootorrataste hulgi- ja jaemüügi ning hoolduse ettevõtete aastaaruannete kasumiaruande ning bilansi näitajatest arvutatud finantssuhtarve.

Lõplikus valimis on 1367 pankrotistunud ettevõtte andmed. Autor jaotab valimi juhuslikkuse alusel treening- ning kontrollvalimiks, sest muul viisil on keeruline teha järeldusi masinõppimise meetodite tulemuste kohta, kuna tihti on klassifitseerimistäpsus treeningvalimil masinõppimise meetodite puhul 100%, mis võib tähendada aga mudelite niinimetatud ületreenimist ning need ei suuda üldistada kontrollvalimeid.

Selles töös kasutatakse autokaubanduse sektori ettevõtete pankroti prognoosimisel 3 meetodit. Esimeseks on logistiline regressioon, mida töö autor kasutas oma bakalaureusetöös Eesi kaubandussektori ettevõtete pankrotistumise prognoosimise mudeli loomisel. Autori poolt varasemalt mitte kasutatud on selles magistrیتöös masinõppimise meetodid: k-lähima naabri meetod ning juhusliku metsa meetod. Esimene neist on oma olemuselt nn “musta kasti” meetod, mille puhul mudeli väljundiks on binaarne hinnang valimis olevate ettevõtete kohta ning juhusliku metsa meetodi puhul on väljundiks rohkem tagasisidet: nii muutujate prognoosivõime, eeldatava klassifikatsioonimäära kui ka muude kasutatavat valimit kirjeldavate statistikute kohta.

Viimaseks ülesandeks seadis autor koostatud mudelite võrdluse ning sektori spetsiifiliste suhtarvude leidmise vastavalt kasutatud meetoditele. Autori eelduste kohaselt võiks parimaid tulemusi saavutada ettevõtete pankrotistumise prognoosimisel masinõppimise meetodeid kasutades kõrvutatuna klassikalise statistilise meetodi ehk antud töös logistilise regressiooniga. Mudelite võrdluse all toob autor lisaks mudelite tundlikkuse, spetsiifilisuse ning üldise klassifikatsioonimäära võrdluse kõrval ka kõikide kasutatud meetodite tugevused.

Käesolevas töös saadud tulemusi on võimalik kasutada autokaubanduse ettevõtete pankrotistumise hindamisel. Antud töö annab suunised, millised finantssuhtarvud on antud sektori ettevõtete pankrotistumise kirjeldamisel olulised. Kuna töös kasutatakse masinõppimise meetodeid, mida on pankrotistumise prognoosimisel vähe kasutatud, siis antud töö näitab kas neid meetodeid on ettevõtete pankrotistumise prognoosimisel võimalik kasutada ning millised on tulemused neid meetodeid rakendades.

1.ETTEVÖTTE PANKROTI PROGNOOSIMISE TEOREETILISED ALUSED

1.1.Ettevõtte pankrot ja selle prognoosimise meetodid

Euroopa riikide seadusandlusele tuginevalt on ettevõtete pankrotistumine riigi spetsiifiline ning riikide vahel võib esineda erinevusi. Autori eesmärgiks järgnevas alapeatükis on selgitada ebaõnnestumise protsessi ning analüüsida, milliseid definitsioone varasemates ettevõtete ebaõnnestumise prognoosimisele keskendunud uurimustes on kasutatud. Käesolevas alapeatükis teeb autor ülevaate seni ettevõtete pankrotistumise prognoosimises enimkasutatud meetoditest kui ka neist, mis vähempopulaarsemad, ent rakendatakse antud uurimustöös. Esmalt pankrotistumise ning ebaõnnestumise defineerimisest.

Piiriüleste maksejõuetust puudutavate kohtuvaidluste pidamine võib tavapäraselt olla oluliselt keerulisem, komplekssem ning kõikehõlmavam kui tavapäraste maksejõuliste ettevõtete vaheliste piiriüleste tülikäsimuste lahendamine. Kui maksejõuetuse definitsioon on rahvusvaheliselt suhteliselt sarnane, siis riigiti on suhtumine pankrotti äärmiselt mitmekesine, kuna tagajärjed võlgnikega juhtuvast varieeruvad riigiti oluliselt. Pankroti seadusandlust võib kirjeldada kui meta-õigust, kuna see puudutab võlgniku tervet “pärandust” ning lisaks valitseb äärmiselt suur huvi nii erinevate huvigruppide kui ka juriidiliste huvide mõistes. Seetõttu side maksejõuetuse seadusandluse ning avaliku huvi vahel riigis on riigis seotud ning vaatamata paljudele suurtele sarnasustele riikide vahel, erinevad maksejõuetuse seadused ning protseduurid riikide vahel suurel määral. (Ghio 2015: 2–3)

Euroopas on viimaste kümnendite jooksul suurenenud maksejõuetuse kaasuste sagedus ning maht. See avaldab kehtiva seadusandluse kitsaskohti riikide ning ka ühenduse tasandil. EL-i riikide maksejõuetuse seadusandluse juures on konflikt riikidevahelise maksejõuetuse tülikäsimuste puhul suurem kui mistahes muude ettevõtete vaheliste tülikäsimuste lahendamise juures. (Ghio 2015: 3)

Ebaõnnestumise definitsioonid, mida on varasemates pankroti või ebaõnnestumise prognoosimise alastes uurimustöodes kasutatud on suuresti sõltunud kasutatavast andmestikust. (Ropega 2011: 477) Eeskätt andmebaasidest, mis on tavaliselt riiklikud, kuid esineb ka erainitsiatiivil põhinevaid ettevõtete majandusnäitajate andmebaase (Bureau van Dijk'i Amadeusi andmebaas, Inforegister, Äripäeva infopank jne.), sõltub milliseid majandusnäitajaid, staatuseid, kuupäevi need sisaldavad.

Enamikes Ameerika Ühendriike käsitlevates pankroti prognoosimise uuringutes kasutatakse ettevõtte ebaõnnestumise definitsioonina pankroti väljakuulutamist, mis on seletatav, kuna peatükk 11 pankrotiseadustikust võimaldab reorganiseerimise korral ebaõnnestunud ettevõtetel tegevust jätkata. Ka Kanadas on suurematel ettevõtetel võimalik Ettevõtete Krediitorite Kokkuleppe seaduse kaitsel reorganiseerimine pankroti vältimiseks. Seega Kanadas pankroti mõiste kasutamine piiraks oluliselt valimi mahtu ning poleks ebaõnnestumise või raskustes ettevõtete mõistes tegu esindusliku valimiga. Seega kasutasid Boritz ning teised (2007:147) ebaõnnestumise definitsioonina pankrotti, pankrotimenetluse alustamist või likvideerimist.

Enamikes uuringutes kasutatakse kahte *a priori* gruppi ettevõtetest, millest ühed on ebaõnnestunud ning teised aktiivsed. Sõltuvalt autori eelistustest, kohalikest oludest (sh. seadusandlusest, andmebaasidest vms.) võib ebaõnnestumise definitsioon varieeruda. Ebaõnnestumise definitsiooniks võib olla pankroti väljakuulutamine ettevõtte poolt, võlakirja tagasimaksete katkemine, pangalaenu tagasimaksete katkemine, ettevõtte börsilt lahkumine, valitsuse sekkumine lisainvesteeringute näol või ettevõtte likvideerimine. Lisaks on äärmiselt tähtis sündmuse juures, mis valitakse ebaõnnestumiseks, selle toimumise kuupäev. Ettevõtete pankrotistumise prognoosimise

järelduste kvaliteet sõltub enamasti sellest, kui varasena on määratud ebaõnnestumise täpne hetk või millist sündmust kasutada ebaõnnestumisena. Aktiivsete ettevõtete andmestik on nõrgem, kuna nende kohta võib öelda, et need kindlasti olid toimivad ettevõtted vaid sel ja ainult sel hetkel, millest need andmed pärinevad. See tähendab, et pole võimalik enamasti väita, kas antud ettevõtted olid tegutsevad ka järgnevatel perioodidel. Sellest tulenevalt on leitud, et paljud teist tüüpi vead ebaõnnestumise prognoosimise mudelites on tingitud neist ettevõtetest, mis järgnevatel ajaperioodidel ebaõnnestuda võivad. (Altman, Narayanan 1997: 2-3)

Tabel 1.1 Ebaõnnestumise definitsioonid eelnevates kaubandussektori spetsiifilistes pankrotistumise prognoosimise uurimistöödes (EÕ - ebaõnnestumine)

Autor	Termin	Definitsioon
Bhargava et al. (1998)	pankrot	Pankrotimenetluse alustamine Ameerika Ühendriikide pankrotiseadustiku peatükkide 11 või 7 alusel.
McGurr, Devaney (1998)	pankrot	Pankrotimenetluse alustamine
Shirata (1998)	EÕ	Juriidiline definitsioon. Kaasatud ka raskustesse sattunud ettevõtete likvideerimine ühinemise tõttu.
He (2006)	pankrot	Pankrotimenetluse alustamine
Lukason (2006)	pankrot	Pankrot ja likvideerimine netovarade nõutele mittevastavuse tõttu
Lee, Choi (2013)	pankrot	Seaduse alusel pankrotistumine
Paal (2013)	pankrot	Kohtu poolt on välja kuulutatud pankrot

Allikas: autori koostatud.

Autori poolt on kokku pandud varasemas kaubandussektori ettevõtete ebaõnnestumise prognoosimisele keskendunud uurimustöodes kasutatud ebaõnnestumise ja pankrotistumise definitsioonide loetelu. Enamasti on kasutatud kaubandussektori põhistes uurimustöodes mõistet pankrot. Vaid He (2006) Jaapani ettevõtete alusel koostatud uurimustöös on tegu ebaõnnestumisega, sest lisaks pankrotistunud ettevõtetele on siin kaasatud ka ebaõnnestunud ettevõtted, mille puhul on teostatud ettevõtete ühinemine.

Ettevõtete ebaõnnestumise ärahoidmiseks tuleb paremini mõista nende ebaõnnestumist ning selleni viivaid põhjuseid, et hinnata, milliste näitajate abil on võimalik seda prognoosida. Ettevõtete ebaõnnestumine pole enamasti akuutne (kuigi esineb ka akuutset pankrotti, mida on peaaegu võimatu prognoosida) sündmus, vaid pigem

dünaamiline protsess (Ropega 2011: 477). Esineb erinevaid käsitusi, kuidas ettevõtteid jõuavad pankrotistumiseni ning millised need protsessid täpsemalt välja näevad.

Ettevõtte “tervis” sõltub võimest teenida kasumit mis peaks tagama piisava maksevõime. Ettevõtte “haigestumine” tähendab selle nõrgestumist hetkeni, mil see muutub maksejõuetuks. Ettevõtte ebaõnnestumine saab võimalikuks kasumlikkuse minetamisel, mille tagajärjena toimub maksevõime vähenemine kuni maksejõuetuseni. See protsess ei ole ainult omane alustavatele ettevõtetele, vaid see võib juhtuda ka börsil kaubeldavate ettevõtetega. Tuginedes Ooghe ja Waevaert’i (2004) pankroti põhjuste liigitusele, võib need jagada viieks. Need on järgmised: üldine keskkond (majandustsüklid, tehnoloogia, välisriigid, poliitiline keskkond ning sotsiaalsed tegurid), vahetu keskkond (kliendid, hankijad, konkurendid, pangad ja krediitiasutused), juhtide/ettevõtjate karakteristikud (motivatsioon, oskused ning isikuomadused), ettevõtte kurss (strateegia ja investeeringud, turundusstrateegia, operatiivsus, personal, finants ja administratsioon, ettevõtte üldjuhtimine), ettevõtte karakteristikud (suurus, arengufaas, tööstusharu ja paindlikkus). Pankrotistumise protsessid on Ooghe ja De Prijcker (2008) jaganud neljaks: edutute *start-up*’ide ebaõnnestumine, ambitsioonikate kasvuettevõtete ebaõnnestumine, väga kiire kasvuga ettevõtete ebaõnnestumine ja üksikõiksete ennast tõestanud ettevõtete ebaõnnestumine. Nende ebaõnnestumise tüüpide puhul on võimalik ettevõtte finantsnäitajatest leida märke halvenevast ärilisest suutlikkusest (Wu 2009: 2372). Seega tuleb finantsnäitajate põhjal arvutada suhtarvud, mille põhjal on võimalik võrrelda ettevõtteid omavahel ning seeläbi hinnata nende ebaõnnestumist.

Praegune ettevõtete ebaõnnestumise definitsioon põhineb rahalise maksejõuetuse kontseptsioonil, millest lähtuvalt on see tihedalt seotud ettevõtte lühiajalise maksevõimega. Dünaamilisi likviidsuse aspekte on võimalik võtta arvesse rahavoo näitajate abil. Staatilisi aspekte mõõdetakse lühiajalise maksevõime kordaja abil. (Laitinen 1991: 656)

Ettevõtte ebaõnnestumise defineerimine on vaid üks osa antud magistritööst. Autor tutvustab varasemalt kasutatud pankrotistumise prognoosimise meetodeid. Eesmärgiks antud valdkonnas on võimalikult täpselt klassifitseerida ettevõtteid, jagades need kas edukateks või ebaedukateks (pankrot/ebaõnnestumine). Ebaõnnestumise prognoosimise mudelid põhinevad enamasti hinnangutel, mis tuginevad aktiivsete ja ebaõnnestunud ettevõtete võrdlusele. Seesuguste mudelite valiidsust hinnatakse klassifitseerimistäpsuse alusel. (Kamath, He 2006: 99) Teatud uuringud on seadnud eesmärgiks võimalikult varase pankrotistumise prognoosimise. Mida varasem on finantsandmete pärimine enne pankrotistumist, seda kesisemaid tulemusi mudelid annavad. Pankrotistumise protsess on üldjuhul mõne aasta pikkune, kuid esineb ka erandeid. Näitajate järkjärguline degradeerumine sõltub ka sellest, millised on mudelitesse kaasatud ettevõtete pankrotistumise protsessid (Stahlman 2015: 77...80).

Pankrotistumise prognoosimiseks on alates 1930. aastast kasutatud mitmeid finantssuhtarvude analüüsil põhinevaid meetodeid. Kumar ning Ravi (2007) jaotavad ebaõnnestumise prognoosimise meetodid kahte laia kategooriasse, milleks on statistilised ning intelligentsete meetodid. Statistilisteks meetoditeks on lineaarne analüüs, mitmene diskriminantanalüüs, ruutdiskriminant- ja faktoranalüüs ning, logit- ja probit mudel. Intelligentseteks meetoditeks loetakse närvivõrkude meetodeid, otsustuspuud, kaasustepõhist arutlemist (*case-based reasoning*), evolutsioonilisi lähenemisi (*evolutionary approaches*), karedate hulki (*rough sets*), hübriid-intelligentseid süsteeme (*hybrid intelligent systems*), lineaarset programmeerimist, andmeradade analüüsi (*envelopment analysis*), ruutprogrammeerimist (*quadratic programming*), tugivektormasinaid (*support vector machine*) ning hägusloogikal põhinevaid süsteeme (*fuzzy logic techniques*). (Kumar, Ravi 2007: 3) Uuteks meetoditeks on: ekspertsüsteemid, närvivõrkude mudelid, karedate hulga teooria (*rough set theory*) ja geneetiline programmeerimine. Suund on järjest keerukamate mudelite poole ning ühe muutujaga mudelitest on jõutud vaheetapina mitme muutujaga statistiliste meetoditeni (diskriminant-, regressiooni- ning lineaarsed diskriminantanalüüsid). Järjest enam on hakatud ettevõtete ebaõnnestumise prognoosimisel kasutama tehisintellekti ning masinõppimisel põhinevaid süsteeme ning mitmed uuringud on näidanud nende paremat

klassifitseerimistäpsust, võrreldes traditsiooniliste statistiliste mudelitega. (Wu 2009: 2372) Antud jaotus on toodud täpsemalt tabelis 1.2 ning seal on toodud ka vastavate meetodite keskmised klassifikatsioonitäpsused. Lisaks tuuakse seni kasutatud mudeleid tutvustavas osas ka need mudelid, mida autor antud töö puhul analüüsiks kasutab.

Tabel 1.2 Ettevõtete pankroti prognoosimise meetodite kategooriad, analüüsi meetodid ja keskmine ennustumäär varasemate uuringute puhul

Kategooriad	Meetodid	Keskmine ennustumäär
Klassikalised statistilised ebaedu prognoosimise meetodid	Ühemõõtmeline analüüs	81,1%
	Mitmene diskriminant analüüs	85,1%
	Lineaarne tõenäosusmudel	80,5%
	Logit mudel	86,7%
	Probit mudel	88,9%
Intelligentsed meetodid ettevõtete pankrotistumise prognoosimiseks	Elukestusmudelid	-
	Otsustuspuu meetod	86,4%
	Närvivõrkude meetod	87,4%
	Multi-logit	-
	CUSUM mudel	84,0%
	Kaose teorial põhinev mudel	-
	Kaasustel põhinev arutlemine (CBR)	83,5%
	K-lähima naabri meetod	-
	Juhuslik mets	-

Allikas: (Aziz, Dar 2006: 19–22, 27; Balcaen, Ooghe 2004: 2–12; Balcaen, Ooghe 2006: 65–71; Kumar, Ravi 2007: 3); autori koostatud ja täiendatud.

Autor toob edasises tekstis kronoloogilises järjekorras enimkasutatud pankrotistumise prognoosimise meetodid. Analüüs kasutades ühte muutujat (*univariate analysis*), oli peamiseks ettevõtete finantsseisundi hindamismeetodiks alates 20. sajandi kolmekümnendatest aastatest ning on kasutuses ettevõtete krediitdivõimekuse ülevaatlikuks ja kiireks hindamiseks praegugi. Interpreteerimise ja rakendamise lihtsus on see, miks kasutatakse ainult ühe muutuja põhjal otsuste tegemist ka praegugi - see võimaldab anda kiire ülevaate ettevõtte seisundist, võrreldes sarnaste (samasse harusse kuuluvate) ettevõtetega. Teatud juhtudel võivad seeläbi antavad hinnangud olla ebatäpsed ja probleeme valmistab ka I liiki vea suhteliselt suur esinemine. Teatud muutujate puhul on niiviisi võimalik kuni viis aastat varem hinnata nende võimalikku pankrotistumist ning ebaedu (Beaver 1966: 91).

Üks esimesi uuringuid, mis käsitleb pankrotistumise prognoosimist finantsnäitajate põhjal (põhiliselt finantssuhtarvudel), on teostatud Bureau of Business Research (1930)

poolt. See uuring võrdles finantsuhtarvude põhjal ebaõnnestunud ning aktiivseid ettevõtteid ning selle uuringu tulemuseks on finantssuhtarvude põhine võimalus eristada ettevõtete ebaõnnestumist ehk kahte erinevasse gruppi (so. pankrotistunud ja aktiivsed) kuuluvate ettevõtete suhtarvude keskmiste tasemete olulist erinevust. (Balcaen, Ooghe 2006: 65)

Ühemõõtmeliste mudelite puhul on suurimaks probleemiks, mis seotud nende püstitusega, nende võimetus anda hinnanguid muutujate koosmõjusid arvesse võttes. Nendes pankroti prognoosimise mudelites leitakse kõigile kaasatud finantssuhtarvudele optimaalsed murdepunktid, mille murdepunktist suurem või väiksem väärtus määrab, millisesse ettevõtete gruppi need kuuluvad: aktiivsete või ebaedukate hulka. Finantssuhtarvu põhiselt erineb, kas antud väärtuse suurenemine tähendab pigem ettevõtte pankrotistumise tõenäosuse vähenemist või suurenemist. Seeläbi saab suhtarvude alusel anda hinnanguid ettevõtete võimaliku käekäigu kohta tulevikus. Majandussektorite lõikes võivad prognoosivõime poolest olulised suhtarvud olla erinevad. Viimane on ka selle uurimustöö koostamise üheks põhjuseks, et teada saada, millised finantssuhtarvud on kaubanduse- ning täpsemalt autokaubanduse sektori jaoks spetsiifiliselt olulised.

Tabel 1.3 Tähtsamate ühemõõtmeliste mudelite autorid ning nende põhjal olulised suhtarvud ning töösse kaasatud suhtarvud

Koostaja	Aasta	Olulised / kaasatud finantssuhtarvud
Bureau of Business Research	1930	8/24
Merwin	1942	3/teadmata
Beaver	1966	6/30

Allikas: (Paal 2013: 14).

Ühe muutuja põhised uuringud ning nendes kasutatud suhtarvude nii oluliste kui ka kõikide analüüsi kaasatud suhtarvude hulk on toodud tabelis 1.3. Need uuringud andsid tõuke mitmemõõtmeliste mudelite rakendamiseks ning Beaver (1966: 100) soovitas edaspidi kasutada keerukamaid ehk mitmemõõtmelisi statistilisi analüüsimeetodeid, mis võimaldaksid hinnata sõltumatute muutujate koosmõjusid. Ühemõõtmeliste finantsuhtarvude senine analüüs näitas, et finantsuhtarvud on üks võimalik ja suhteliselt

lihtsasti kasutatav metoodika ettevõtte hetkeseisundi ning ka tulevikuperspektiivide hindamiseks. See andis alust arvata, et nende kasutamine keerukamates mudelites võib anda veel paremaid tulemusi (Ibid.: 100). Sektorit hästi tundev spetsialist võib öelda suhtarve hinnates, millises seisus ettevõtte parajasti on - selleks pole vaja keerukaid mitmemuutujaga mudeleid.

Beaveri (1966) aasta uurimuse põhiliseks järelduseks oli ühe muutujaga (*univariate*) analüüsi väike klassifikatsioonitäpsus ettevõtete pankrotistumise prognoosimisel muutujate võimalike koosmõjude mitteamestamise tõttu. Sellest lähtuvalt andis ta soovitus kasutada mõnda statistilist analüüsimeetodit, mis võimaldaks muutujate koosmõjude arvestamist. Kaks aastat hiljem avaldas Altman (1968) uurimuse: *“Financial ratios, discriminant analysis and the prediction of corporate bankruptcy”*, millest sai valdkonna teedraja teadustöö. Seniajani kasutatakse mitmemõõtmelist diskriminantanalüüsi pankrotistumise prognoosimisel. Sellest alates on järjest keerukamaid statistilisi meetodeid hakatud valdkonnas kasutama.

Vaatamata pankrotistumise prognoosimiseks oluliselt paremate statistiliste meetodite väljatöötamisele on siiani enimkasutatavaks pankrotistumise prognoosimise meetodiks mitmene diskriminantanalüüs (*MDA*) ning seejärel logistiline regressioon, mida tuntakse ka terminiga logit mudel. Pigem kasutatakse teaduskirjanduses viimast nimekuju (Adnan Aziz, Dar 2006: 26).

Eelnevalt loetletud meetodite suur populaarsus on seletatav nende püstitamise ning tulemuste interpreteerimise suhtelise lihtsusega. Saavutatavad õigesti klassifitseerimise, tundlikkuse ning spetsiifilisuse määrad on üldjuhul suhteliselt kõrged. Mitmese diskriminantanalüüsi puhul on tulemuste interpreteerimine mõnevõrra lihtsam kui logistilise regressiooni puhul. Mudelite tulemuste interpreteerimine on ka statistika spetsiifikat mittevõlaldavale indiviidile küllaltki intuitiivne, kuna nende mudelite tavapärane ülesehitus on piisavalt lihtne ja seetõttu ka üheselt mõistetav. Tulemuste interpreteerimiseks saab kasutada mitmese diskriminantanalüüsi korral sõltumatute muutujate ees olevaid mudeli kordajaid ja logit mudeli puhul saab selleks arvutada

marginaalsed efektid. Mudeli kordajad (*MDA*) kui ka marginaalsed efektid (logit mudel) võimaldavad analüüsida mudeli skoori tundlikkust sõltuvate muutujate väärtuse ühikulistele muutustele. Mitmese diskriminantanalüüsi puhul on võimalik analüüsida mudeli skoori muutumist, kuid logit mudeli puhul annavad marginaalsed efektid võimaluse hinnata pankrotistumise tõenäosuse muutust sõltumatute muutujate ühikuliste muutuste korral. (Balcaen, Ooghe 2006: 66 – 70)

Diskriminantanalüüsi (*multiple discriminant analysis*) puhul luuakse sõltumatuid tunnuseid kasutades mudel, mis koosneb sõltumatutest muutujatest ning nendele arvutatud diskriminantfunktsiooni kordajatest ning vabaliikmest. Iga vaatluse korral on antud lineaarse funktsiooni väärtus hinnanguks, mis kirjeldab sõltuvat muutujat. Selle meetodi puhul on võimalik klassifitseerida vaatlused kuuluvaks ühte mitmest eeldusena valitud grupist. Seega annab meetod võimaluse objekti mitme eri tunnuse põhjal teisendada üheks hinnanguks. Mitmese diskriminantanalüüsi diskriminantfunktsioon avaldub järgneval kujul (Ooghe et al. 2009: 9, Laitinen et al. 1996: 3):

$$(1.1) \quad D_i = d_0 + d_1 X_{i1} + d_2 X_{i2} + \dots + d_n X_{in},$$

kus D_i – i-nda ettevõtte Z-hinnang vahemikus $-\infty \dots +\infty$;

X_i – mudeli sõltumatud muutujad;

d_i – diskriminantfunktsiooni kordajad.

Altmani 1968. aasta Z-skoori mudeli järgselt, mis põhines diskriminantanalüüsil, on tekkinud nn ebaõnnestumise prognoosimise buum ning seejärel on kasutatud mitmeid erinevaid meetodeid sealhulgas, logit mudelit, probit mudelit ja lineaarset programmeerimist. Eelnevalt loetletud mudelite, võib ka öelda konventsionaalsete statistiliste meetodite suurimaks probleemiks või kitsaskohaks on andmestikule esitatavad nõuded: lineaarsus, normaaljaotusele vastavus ning sõltumatus muutujate vahel. Nende eelduste ehk siis statistiliste eelduste vältimiseks on tekkinud vajadus uute meetodite järele.

Ettevõtete pankrotistumise prognoosimise mudelite puhul kasutatakse enamasti kahte gruppi, milleks on ebaõnnestunud-ning aktiivsed ettevõtted. Kahe valitud *a priori* grupi põhjal luuakse mudel. Lõpptulemuseks on mudel, mille alusel saab arvutada iga suvalise ettevõtte (mille puhul on olemas vajalikud sisendandmed) põhjal hinnangu, mille alusel saab selle finantsolukorda hinnata. Selleks on diskriminantfunktsiooni väärtus selle konkreetse ettevõtte korral. (Altman 1968: 591) Murdepunkti valik nendele hinnangutele, mis saadud funktsiooni alusel arvutatakse, toimub lähtuvalt uuringu läbiviija eesmärgist. (Daubie et al. 2002: 690). Murdepunkti valiku kriteeriumiteks on näiteks järgnevad üksteist välistavad tingimused: valesti klassifitseerimise määra vähendamine, I liiki vea esinemise vähendamine, II liiki vea esinemise vähendamine (Hsieh 1993: 461) või mõlema grupi klassifikatsioonimäära võrdsustamine (Ohlson 1980: 125).

Lineaarse tõenäosusmudeli (*linear probability model*) abil on võimalik hinnata ettevõtete pankrotistumise tõenäosust. Regressioonimudeli sõltuv muutuja on binaarne, mis annab kaks võimalust: pankrotistuv või aktiivne. (Dimitras et al. 1995: 503) Lineaarne tõenäosusmudel on üks tingliku tõenäosuse hindamise mudelitest (Balcaen, Ooghe 2006: 68). Lineaarse tõenäosusmudeli puhul on määramispiirkonnaks 0...1. Tingliku tõenäosuse meetoditest on enim kasutatud logit- ja probit mudelid, millest esimene on ettevõtete pankrotistumise mudelite koostamisel üks populaarsemaid meetodeid.

Lineaarse tõenäosusmudeli koefitsientide arvutamiseks kasutatakse ühe võimalusena vähimruutude meetodit. Meetodi alusel arvutatakse ettevõtetele nende seisundit hindavad skoorid. Skoore saab tõlgendada ettevõtete ebaõnnestumise tõenäosustena (Theodossiou 1991: 701). Lineaarse tõenäosusmudeli puhul võib juhtuda, et mudeli muutumispiirkond on laiem kui 0...1 ning seetõttu ei ole võimalik sellest välja jäävaid väärtusi tõlgendada kui tõenäosusskoori. Mudeli teisendatud kuju, mille alusel hinnatakse ettevõtete ebaõnnestumise tõenäosust, on toodud järgnevalt (Dimitras et al. 1995: 503):

$$(1.2) \quad P_i = a_0 + a_1 x_{i1} + a_2 x_{i2} + a_3 x_{i3} + \dots + a_n x_{in},$$

kus, P_i — i-nda ettevõtte ebaõnnestumise tõenäosus;

a_i — vähimruutude meetodil arvatatud mudeli hinnangud;

x_i — ettevõtete põhised sõltumatud muutujad.

Ühte tingliku tõenäosuse (*conditional probability*) hindamise mudelit ehk logit mudelit kasutas ettevõtte pankroti prognoosimiseks esimesena Ohlson (1980). Tingliku tõenäosuse hindamise mudelid võimaldavad ettevõtte karakteristikute ehk peamiselt finantsuhtarvude ning muude kirjeldavate tunnuste alusel hinnata ühe skoorina ettevõtte ebaõnnestumise tõenäosust. Arvatav logit skoor ei ole absoluutse kindlusega tõenäosuslik hinnang ettevõtte pankrotistumiseks, vaid tegu on logistilise hinnanguga mis on arvatatud a priori binaarse sõltuva muutuja ning uurija valitud sõltumatute muutujate vahel. Mudeli koefitsientide arvutamine toimub logaritmilise tõepära funktsiooni maksimeerimisel (Dimitras et al. 1995: 504). Logit mudeli väljundiks on ettevõtete pankrotistumise tinglik tõenäosus ehk logit skoor, mille muutumispiirkonnaks on 0...1. (Dimitras et al. 1995: 504). Logit mudeli abil on võimalik mitmemõõtelise ehk siis vastavalt n muutuja koosmõju hindamine ühemõõtmelise logit skoorina.

Logit mudeli üldkuju ning muutjate seletused on avaldatud alljärgnevalt (Ooghe et al. 2009: 9, Laitinen et al. 1996: 3):

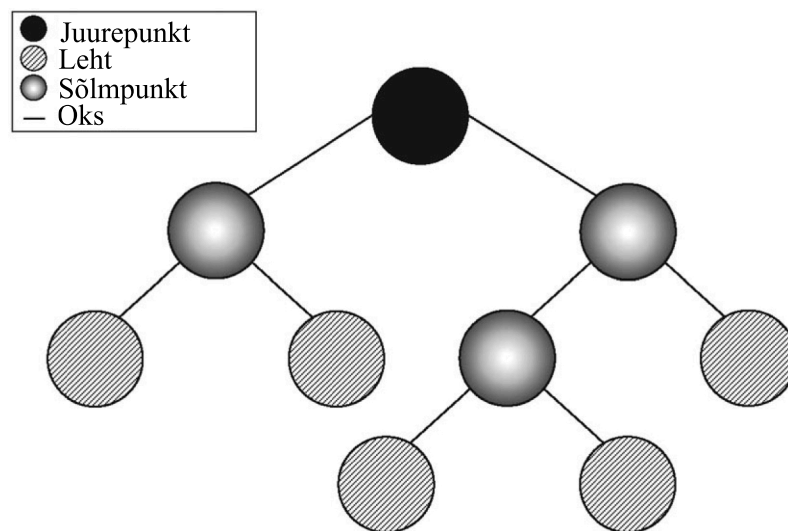
$$(1.3) \quad L = \frac{1}{1 + e^{-(b_0 + b_1 V_1 + b_2 V_2 + \dots + b_n V_n)}}$$

kus, L_i — logit skoor vahemikus 0...1 ehk sisuliselt pankrotistumise tõenäosus;

V_i — mudeli sõltumatud muutujad;

b_i — vabaliige (0) ning koefitsiendid.

Mudeli hinnang peaks olema ideaaljuhul kas null või üks, vastavalt siis kas pankrotistumise tõenäosus on null ehk tegemist on aktiivse ettevõttega või üks pankrotistuva ettevõtte puhul (vastupidine kodeering ehk aktiivne ettevõtte “1” ning pankrot “0” muudab mudelilt saadavate hinnangute polaarsust). Tegelikuses võib see skoor antud piirides varieeruda ning L-väärtust võib nimetada ettevõtte pankrotistumise tõenäosuseks. Ideaalis peaks olema murdepunkt väärtusega 0,5. Mudeli püstitusest sõltuvalt võib see murdepunkt erineda, olles vastavalt suurem või väiksem. Mudeli murdepunkti valikul võib seada mitmeid erinevaid kriteeriumeid: vähendada I liiki viga, vähendada II liiki viga, suurendada õigesti klassifitseerimise määra või muu mudeli koostaja poolt seatud tingimus. (Ohlson 1980: 124–130) Mudeli murdepunkti valik sõltub seega autori poolt seatud eesmärkidest ning võib ka ühe uurimuse raames olla mitmeti defineeritud ning esitatud.



Joonis 1.1 Binaarse otsustuspuu baasstruktuur (autori koostatud; Bhattacharya et al. 2010: 539 põhjal).

Logit mudeli rakendamisel on suurimaks probleemiksmultikollineaarsuse esinemise võimalik oht muutujate komplektis. Tugevasti korreleerunud muutujate kaasamist ühe mudeli hindamisel tuleks vältida (Balcaen, Ooghe 2004: 20). Mudeli parameetrite sobivuse hindamiseks on võimalik kasutada näiteks tõepära suhtarvu ja Waldi statistikut. Mudeli headuse hindamiseks on näiteks Hosmer-Lemeshow test. (Harrel 2001: 228-231) Logit mudeli eeliseks on tema kasutamise suhteline lihtsus ning mudeli parameetrite alusel hinnangute tegemine. See annab võimaluse hinnata ettevõtte

finantsseisundit ning pankrotistumise tõenäosuslikku ohtu. Antud töös ei too autor välja probit mudelit, kuna see on küllaltki sarnane logit mudeliga ning pankroti prognoosimisel võrreldes viimati nimetatule oluliselt vähem kasutust leidnud.

Järgnevalt tutvustab autor alternatiivseid pankroti prognoosimise meetodeid, millest esimesena tuleb käsitleda otsustuspuu meetodil põhinev mudel, mille põhijooneks on gruppide alusel optimaalse otsustuspuu genereerimine, viimase alusel on võimalik objekte hilisemalt klassifitseerida ühte või teise kuuluvaks. Antud juhul tähendab see ettevõtete klassifitseerimist aktiivseteks või pankrotistuvateks. Otsustuspuu on binaarne ehk igast tasandist edasiliikumiseks on kaks võimalust. Otsustuspuu (vt. joonis 1.1) luuakse varem defineeritud gruppide põhjal ning nende alusel genereeritakse hulk sõlmpunkte, millel on üksteist välistavad jaotustingimused. Sõlmpunktist edasi liikumine toimub mööda niinimetatud oksa vastavalt järgmisesse sõlmpunkti või lehele ehk lõpp punkti. Viimastes sõlmpunktides enne lehti toimub lõplik klassifitseerimine ning lehed on binaarse klassifikatsiooni lõpp-punktid. (Bhattacharya et al. 2010: 539)

Otsustuspuude kasutamisel on mitmeid eeliseid, võrreldes statistiliste meetoditega. Üheks eeliseks on muutujatele mitteseatav kindla statistilise jaotusele vastamise nõue (näiteks: normaaljaotus, unimodaalsus). Teiseks mudelisse kaasatavate muutujate vahel võib esineda kõrge korrelatsioon või multikollineaarsus. Selle meetodi puhul üheks eeliseks visuaalne ja lihtsasti mõistetav kuju. (Daubie et al. 2002: 687)

Üks otsustuspuu võimalikest variantidest näiteks on CHAID (*CHi-squared Automatic Interaction Detection*) tüüpi regressioonipuud. CHAID on sarnane CART regressioonipuule, kuna mõlema puhul genereeritakse otsustuspuu, kuid nad erinevad üksteisest viisi osas, kuidas seda tehakse. Kui CART meetodi puhul on valiku kriteeriumiks entroopia või Gini meetrika, siis CHAID-i puhul on optimaalse jaotuspunkti valikukriteeriumiks hii-ruut test. CHAID meetodi erinevuseks on ka see, et selle puhul ei pea olema tulemuseks binaarne otsustuspuu, vaid otsustuspuu võib jaotada rohkematesse gruppidesse. Meetodi puhul võivad sõltumatud muutujad olla nii

kategoorilised kui ka pidevad, kuid viimase variandi puhul pidev muutuja kategoriseeritakse. (Zheng, Yanhui 2007: 2; Andreica et al. 2013: 299)

Juhuslik mets (*random forest*) on meetod, mis juurutati Breimani poolt 2000. aastal. See on masinõppimise meetod, mille puhul klassifitseerimine toimub “n” arvu otsustuspuude põhjal. Otsustuspuude arvu valik toimub vastavalt tulemuste stabiliseerumisele puude lisandudes. Tegemist on mitteparameetrilise meetodiga, mis tähendab, et andmed ei pea vastama ühelegi klassikaliste statistiliste meetodite poolt nõutud eeldusele (normaaljaotus, multikollineaarsuse puudumine). Andmestik võib olla väga kiivas jaotusega või esineda näiteks mitu moodi ehk multimodaalsus. Lisaks võivad olla kas järjestatud või järjestamata struktuuriga kategoorilised muutujad. Meetod on kasutatav ka siis, kui muutujate arv on suurem kui vaatluste arv. Mudelil ei ole kalduvust ületreenimiseks (*overfitting*) ehk treeningvalimi liiga täpseks järgimiseks, mille puhul kontrollvalimite puhul mudeli klassifikatsioonimäär drastiliselt kahaneks. Mudelite klassifikatsioonitäpsuse hindamiseks kasutatakse OOB (*out of bag* ehk nn kotist väljavõtmist) klassifikatsiooni viga. (Liaw, Wiener 2002: 18)

K-lähima naabri meetod (*k-nearest-neighbor discriminant analysis*) on mitteparameetiline ehk jaotuse vaba meetod, mida kasutatakse vaatluste klassifitseerimiseks gruppidesse pidevate tunnuste alusel. Selle meetodi kasutamiseks pole vaja eeldusena normaaljaotust ning andmed ei pea vastama ka muudele parameetritele, millele diskriminant- ning logistilise regressiooni puhul peavad vastama. Vahemaa ehk kaugus kahe vaatluse vahel $d(x,y)$ on tavapäraselt defineeritud mingi funktsiooni alusel (Eukleidese distant, Mahalanobise distant, jne). Kasutades lähima naabri otsustusreeglit määratakse igale vaatlusele grupp, kuhu enamus tema k-lähimatest naabritest kuulub ning selle alusel toimub edasine jaotamine. (Kiang 2003: 445-446)

Kumulatiivse summa mudel ehk CUSUM on dünaamiline edasiarendus mitmesest diskriminantanalüüsist, kuna selle abil on võimalik paremini arvesse võtta ettevõtte majandusnäitajate muutust ajas. Selle mudeli abil on võimalik vahet teha muutujate

autokorrelatsioonist põhjustatud hetkelistel muutustel ning pikemaajalistel muutustel, mis on tingitud finantsprobleemidest. Järjestikkuse protseduuri abil on võimalik hinnata ettevõtte näitajate ühisjaotuse liikumist headelt näitajatelt halvemuse poole. Finantsnäitajate ühisjaotuse nihet peetakse tendentsiks ettevõtte võimaliku pankroti poole liikumisest. (Balcaen, Ooghe 2004: 14) Meetodi eeliseks teiste ees on ettevõtte finantsolukorra hetkeseisundi ning mineviku hindamine. Sellel on lühike mälu ettevõtte hea käekäigu üle, kuid pikk mälu ettevõtte halbade näitajate suhtes. Viimasest tuleneb, et mudel arvestab mitut perioodi, ja ebaõnnestumise kirjeldamisele aitab kaasa ka mudeli tundlikkus halbadele näitajatele ning selle suhteline vähemtundlikkus head käekäiku hindavatele näitajatele perioodide lõikes. (Theodossiou 1993: 448)

Kaasustel põhineva arutluse süsteemide (*Case-Based Reasoning*) aluseks on varasematele juhtumitele tuginev andmebaas ning hinnangute andmine toimub vastavalt üksikjuhtumile. Oluline selle meetodi juures on piisavalt suur andmebaas varasematest juhtumitest, kusjuures varasemate juhtumite kogumit on võimalik ajas laiendada ning kitsenduseks ei ole selle meetodi puhul teatud näitajate puudumine ettevõtetel või ka teatud suvalisest hetkest uute näitajate lisamine kogumisse. See on üks meetodi eelistest ja seda võib nimetada niinimetatud iseõppivaks süsteemiks, mis eelnevate näidete puhul ajas paremaks muutub. (Han et al. 1997: 98) Sarnasuse alusel on võimalik anda hinnanguid ka muudele huvipakkuvatele probleemidele, kasutades juba koostatud andmebaasi, kuid seda juhul, kui andmebaasis on piisavalt informatsiooni (Aziz, Dar 2006: 21).

Käesolevaga on antud ülevaade seni pankroti prognoosimisel kasutatud meetoditest. Sellesse ei kuulu kindlasti absoluutselt kõik meetodid, mida selles vallas kasutatud on, kuid need peaksid andma hea ülevaate, milliseid erinevaid võimalusi saab kasutada ettevõtete finants- või muude näitajate põhjal võimaliku ebaedu või jätkusuutlikkuse hindamisel. Mitmetes artiklites on kasutatud ka eelnevalt loetletud meetodite hübriide ehk loodud mitmel meetodil põhinevaid hübriidseid mudeleid. Osa siintoodud meetodeid on kindlasti võimalik kasutada ka muuks kui vaid binaarseks klassifitseerimiseks.

1.2.Kaubandussektori pankroti prognoosimisel kasutatavad muutujad

Ebaõnnestumise prognoosimise mudelite koostamisel on kasutatavate muutujate kõige esmaseks valikutingimuseks olnud nende kättesaadavus. Antud töös on olnud autoril kasutada üleeuroopaline *Amadeus*’e andmebaas ning selles töös kasutatavate võimalikele muutujate valikule on see piiranguks, kuna sisaldab kindlat komplekti finantsnäitajaid. Sellelaadne piirang olnud ka varasemates uuringutes, kuid antud fakt ei leia nendes üldiselt kajastamist. Andmebaasides leiduv võimalik arv erinevaid finantsnäitajaid on aja jooksul pigem kasvavas trendis ning võib eeldada, et mida aeg edasi seda suuremat hulka neist tuletatud erinevaid finantsuhtarve on võimalik kasutada. Viimast lubab oletada nii arvutusvõimsuse kui ka statistiliste ja/või intelligentsete süsteemide areng. See tähendab, et mudelitesse võib kaasata nii palju erinevaid muutujaid, kui andmestiku põhjal on võimalik arvutada. Autor tutvustab järgnevalt muid olulisi valikukriteeriume, mille alusel on selle valdkonna ehk ettevõtete ebaõnnestumise prognoosimise olulistesse uuringutesse muutujaid valitud.

Antud valdkonna uuringutes on valdavalt uuritud tööstusettevõtteid. Kahjuks pole varasemalt uuritud autokaubandusettevõtteid ning suhteliselt vähe ka kaubandusettevõtteid. Järgnevalt toob autor ülevaate, milliseid muutujaid on kasutatud pankroti prognoosimiseks ning kaubandussektori mudelites oluliseks osutunud muutujad tuuakse välja eraldi tabelites. Siinkohal tuleb arvestada NACE rev. 2 G45 sektori (Mootorsõidukite ning mootorrataste hulgi- ja jaemüük ning hooldus) eripäradega. Sellesse sektorisse kuuluvad täpsemalt (Statistical... 2016):

G45.1 - mootorsõidukite müük;

G45.2 - mootorsõidukite hooldus ning parandus;

G45.3 - mootorsõidukite osade ning tarvikute müük;

G45.4 - mootorrataste müük, hooldus, parandus ning nendega seotud osade ja tarvikute müük.

Kuna antud uurimustöös kajastatava sektori puhul on lisaks hulgi- ning jaemüügile tegu ka vastavate sõidukite ehk siis kas mootorsõidukite või mootorrataste hooldusega, siis tuleb tähelepanu eraldi pöörata ka hoolduse ehk siis teeninduse eripäradele. Teenindava poole jaoks olulised suhtarvud erinevad eelduslikult kaubanduse suhtarvudest kasvõi näiteks tööjõukulude osakaalu poolest, millest esimese puhul peaksid tööjõukulud suuremad olema. Võib ka eeldada, et nende kahe üksteisest erineva tegevusvaldkonna puhul on rentaabluste näitajad erinevad. Autor toob antud alapeatükis välja kaubandussektori mudelites oluliseks osutunud muutujad ning annab ülevaate finantssuhtarvude kategooriatest.

Ebaõnnestumise prognoosimise puhul on peamiseks huvisuunaks koostada mudel, mis kajastaks suhet finantssuhtarvude ning pankrotistumise vahel. Vähemoluliseks ei saa pidada selle mudeli uuesti rakendamise võimalust. Algselt lisatakse mudelisse palju erinevaid tunnuseid ehk iga tunnuse lisamisega lisandub mudeli mõistes üks dimensioon ning mudeli koostamiseks kasutatav treeningvalim ja selles leiduv info võib seeläbi osutada ebapiisavaks. Viimast kutsutakse dimensionaalsuse needuseks, mis tähendab, et muutujate arvu lineaarsel kasvamisel vajalik info hulk ehk vaatluste arv peab eksponentsiaalselt kasvama, et tagada analüüsimiseks vajalik info hulk välistamaks lünkasid andmestikus. Seega kujutab tunnuste valik endast probleemi, milliseid ebaolulisi, üleliigseid ning korreleeruvaid tunnuseid klassifikatsioonimudeli täpsuse, lihtsuse ja dimensionaalsuse needuse vältimise saavutamiseks eemaldada. Kõik see võimaldab vähendada arvutusliku võimsuse vajadust ning aitab parandada mudelite kasutamist väljaspool teadust. (Lee, Choi 2013: 2942)

Ühe tuntuma ning enimsiteeritud mudeli ehk Altman'i (1968: 594) Z-skoori arvutamiseks on vaja viit muutujat, milleks on järgmised: käibekapitali suhtarv koguvaradesse, jaotamata kasumi suhtarv koguvaradesse, EBIT-i suhe koguvaradesse, omakapitali turuväärtuse suhe kogu kogu võla raamatupidamis-väärtusesse ning käibe suhtarv koguvaradesse. Antud uuringus kasutas ta mitmest diskriminantanalüüsi. Deakin (1972: 169) seevastu tõi muutjaid juurde ning kasutas neljateist muutujat, mille jaotas nelja gruppi: mittelikviidsete varad suhtarvud, käibevara suhtarvud

koguvardesse, käibevarade suhtarvud lühiajalistesse kohustustesse ning käibevarade suhtarvud käibe mahtu.

Mitmete mudelite loomise puhul on muutujate valikul tuginetud varasematele uuringutele. Siinkohal toob autor valiku uuringutest, mille muutujate valikul on kasutatud eelnevalt kirjeldatud printsiipi. Gilbert *et al.* (1990: 164) valisid enda logit mudeli abil koostatud pankrotimudelisse Casey ja Bartczaki (1985) mudeli 6 suhtarvu ning Altmani (1968) mudelist pärit 5 suhtarvu ning seega kasutasid mudelis 11 erinevat suhtarvu. Samas Beaver (1966) kritiseerib mudelitesse varasemate uuringute põhjal tehtavat muutujate valikut, kuna enamasti nendesse viidatavatesse töödesse on muutujate valik tehtud “kõhutunde” alusel. Muutujate valikul teistest töödest võib tekkida paratamatult oht, et ka need varem kasutatud muutujad võivad olla mitterelevantsed. Ohtu kujutab see, et ettevõtted võivad hakata nende krediitdivõimekuse või pankroti ennustamiseks kasutatavate finantsuhtarvude arvutamiseks kasutatavaid alusnäitajaid manipuleerima ning seega nende kasutamine võib anda oodatust kehvemaid tulemusi. Sama on tõdenud ka Wu ning Ho (1997: 75), et paljud uuringud põhinevad varasemates uuringutes palju kasutamist leidnud finantssuhtarvudel ning ettevõtete juhatused manipuleerivad nende väärtustega, kuna vastavad väärtused on nende tulemustasustamise ja/või ettevõtete omavahelise võrdluse aluseks. Nendeks näitajateks on enamasti ROA, ROE, EVA, käibe suhtarvud, koguvõla suhtarv koguvardesse, intressi kattekordaja, ROI. Seetõttu on ebatraditsioonilised finantsuhtarvud, mille arvutamiseks kasutatakse raskemini manipuleeritavaid finantsnäitajaid, suurema ennustusvõimega.

Autor võttis kokku muutujate valiku kriteeriumid eelnevates uurimustes, kuid sellega seotult jättis avamata põhjused, millised ettevõtte tegevuse suundasid konkreetsed valitud suhtarvud kajastavad. Mitmete uuringute puhul oli valikukriteeriumiks nende eelnev kasutamine akadeemilises kirjanduses (Han *et al.* 1997: 7). Kuid järgnevalt leiavad kajastamist põhjused miks erinevaid suhtarvude kategooriaid või suhtarvud on mudelitesse kaasatud. Esimeseks oluliseks faktoriks suhtarvude kaasamisel mudelitesse on nende roll ebaõnnestumise protsessi kirjeldamisel (Altman 1968, Theodossiou 1991).

Teiseks valikukriteeriumiks võib olla mudelite hindamisel statistilise olulisuse omandanud näitajad. (Skogsvik 1990: 140) Kolmandaks, siinkohal piiravaks faktoriks muutujate mudelisse kaasamisel on nende arvutamiseks vajalike tegurite puudmine andmebaasidest.

Klassikaliste statistiliste meetodite puhul on mudelitesse kaasatud muutujate arv tavapäraselt 4...20 vahel ning pigem eelistatakse meetoditele esitatud statistiliste eelduste täitmiseks kasutada väiksemat muutujate arvu. Üheks põhiliseks põhjuseks muutujate arvu valikul on multikollineaarsuse võimalik esinemine muutujate komplektis. See tuleneb finantssuhtarvude arvutusvalemite, millede puhul võib tihti esineda samu tegureid arvutusvalemite. Nendeks on näiteks suhtarvud koguvaradesse või müügikäibesse. Multikollineaarsus võib esineda ka erinevast perioodist pärit samade näitajate vahel, kuna nende arvutamiseks on kasutatud erinevast perioodist pärit finantsnäitajaid. Sel juhul ettevõtete põhised muutused finantssuhtarvude tasemetes on marginaalsed ning erinevate perioodide finantssuhtarvude vahel on seeläbi ülikõrge korrelatsioon, mis omakorda on multikollineaarsuse põhjustajaks (Han et al. 1997: 7).

Muutujate valikuks ettevõtete ebaõnnestumise prognoosimise mudelitesse pole üldtunnustatud kriteeriumeid, mistõttu kasutatakse teaduskirjanduses mitmeid erinevaid lähenemisviise. (Kahya 1997: 67; Horrigan 1968: 285) Mudelisse võib kaasata näitajaid, kirjeldamaks ettevõtte suurust või näiteks sularaha- või laovarude hulka kui ka finantsnäitajatel põhinevaid suhtarve, mis võimaldavad kirjeldada rentaablust, maksevõimet ja teisi suhtelisi näitajaid (Horrigan 1968: 285). Finantssuhtarvud võetigi kasutusele ettevõtte krediitvõimaluste hindamiseks ning on saanud üheks olulisemaks näitajaks kajastamaks ettevõtte hetkeseisundit krediidiandjatele, krediidiireitingut andvatele ettevõtetele, investoritele ning juhatusele (Beaver 1966: 71).

Pankrotistuva ettevõtte seisundi muutumise sümptomite kindlakstegemine võimaldab leida, milliseid finantsnäitajaid mudelitesse kaasata. Ebaõnnestumise finantsilised põhjused saab jagada üldistatuna kaheks: puudulikkust lühiajalisest maksevõimest ja

pikaajalisest maksevõimest tulenevateks probleemideks. Need mõlemad on võimalik kindlaks teha vastavate finantssuhtarvude alusel. (Laitinen 1995: 434)

Pankroti prognoosimise mudelites saab kasutada ka ettevõtte suurust kajastavaid näitajaid, et eristada erineva suurusega ettevõtteid, mille puhul väikeste pankrotistumise tõenäosus peaks olema autori hinnangul suurem kui suurte ning väljakujunenud ettevõtete pankrotistumine. Ettevõtte tegevust on võimalik kirjeldada ka muude kvalitatiivsete hinnangute abil, näiteks kajastus meedias, juhatuse liikmete vahetumise arv mingi perioodi jooksul. Nimistut võib jätkata, kuid kvalitatiivsete näitajate põhjal ettevõtete pankrotistumise prognoosimine pole antud uurimustöö eesmärk ning autor ei peatu sellel pikemalt.

Andmekaeve on üks alternatiivne võimalus muutujate valikuks. Üks enimkasutatud andmekaeve tööriistu on klassifikatsiooni ja regressioonipuu mudel (CART). Klassifitseerimise uuringu eesmärgiks võib olla leida täpne klassifikaator või avastada pankrotistumise struktuur. CART mudeli puhul toimub muutujate valik ilma ühegi statistilise eelduseta ning seetõttu ei pea eemaldama ühtegi finantsuhtarvu, mis ei vasta meetodi eriti klassikaliste statistiliste meetodite puhul seatud eeldustele. See võimaldab vältida andmekadu. (Shirata 1998: 2-3)

Beaveri (1966: 80) likviidsete varade reservi mudel vaatlleb ettevõtet kui likviidsete varade reservi. Ettevõtte koosneb likviidsetest varadest, mille maht on reserviks. Nende maht suureneb sissetulevatest rahavoogudest ja kahaneb ettevõttest väljaminevate rahavoogudega. Reserv on puhvriks sissetulekute ning väljaminekute varieerumisele. Ettevõtte maksevõimet on defineeritud selle reservi võimaliku otsalõppemise tõenäosusega, mil ettevõtte muutub maksevõimetuks ja sellest tulenevalt ebaõnnestub. (Beaver 1966: 80)

Beaver on ettevõtte neljast finantskomponendist lähtuvalt sõnastanud järgnevad ettevõtte ebaõnnestumise tõenäosuse muutumist kirjeldavad teesid (Ibid. 1966: 80):

1. mida suurem on reserv, seda väiksem on ebaõnnestumise tõenäosus;

2. mida suurem on likviidsete varade puhasvoog põhitegevusest, seda väiksem on ebaõnnestumise tõenäosus;
3. mida suurem on ettevõtte võlgade kogumaht, seda suurem on ebaõnnestumise tõenäosus;
4. mida suurem on rahaline väljaminek põhitegevuse finantseerimiseks, seda suurem on ebaõnnestumise tõenäosus.

Beaveri poolt välja toodud teese on autori arvates oluline arvesse võtta finants-suhtarvude mudelisse valimisel, kuna sellega on arvesse võetud nii reservid, likviidsus, kohustuste suurus ning põhitegevuse finantseerimiseks vajaminev rahavoog. Bilderbeek (1979) omakorda kasutas nelja erinevat suhtarvude kategooriat, kirjeldamaks ettevõtte seisundit: rentaabluse, efektiivsuse, lühiajalise maksevõime ja pikaajalise maksevõime näitajaid. Ettevõtte liigub enne ebaõnnestumist järk-järgult probleemide suunas. Probleemid algavad ettevõtte efektiivsuse kahanemisega, mille tulemusena toimub kasumi vähenemine, kuni ettevõtte muutub kahjumlikuks. Sellest tulenevalt langevad efektiivsuse ja rentaabluse näitajate tasemed. Mitmeaastane kesine rentaablus nõrgestab ettevõtte maksevõimet, mis kajastub maksevõime näitajate halvenemises. Võlausaldajate poolne krediidi pakkumise lõpetamine nõrga maksevõime ning madala kasumlikkuse tõttu põhjustab likviidsete vahendite järsu nõrgenemise ja toimub hetk enne pankrotistumist. (Bilderbeek 1979, viidatud Bilderbeek, Pompe 2005: 849 vahendusel)

Uurijad on pankrotistumist mõjutavaid tegureid mitmeti loetlenud, kuid puudub siiski ühene teoreetiline mõistmine, millised muutujad võiksid oma pankroti ennustusvõimelt olla paremad ja milline võiks olla optimaalne muutujate valik. See takistab pankroti prognoosimise mudelite arengut (Zavgren 1983: 5). Muutujate valik ettevõtete pankroti prognoosimise mudelitesse toimub pigem intuiitselt, toetudes varasematele uuringutele ja ühemõõtmelistele statistilistele meetoditele (t-test, F-test, korrelatsioonanalüüs) (du Jardin 2012: 21).

Klassikaliste statistiliste meetodite puhul kontrollitakse enne mudeli koostamist valimis sisalduvate muutujate statistiliste eelduste täidetust, mis varieeruvad vastavalt

kasutatavale meetodile. Nõuete mittetäidetuse või vähese kirjeldusvõimega muutujad eemaldatakse edasisest analüüsist, mille tulemusel jõutakse statistiliselt kirjeldusvõimelt parima lahendini. Parim lahend kehtib antud valimi korral, mistahes muu valimi korral võib lahend olla teistsugune. (Scott 1981: 321) Muutujate ennustusvõime sõltub mitmemõõtmeliste mudelite puhul ka koosmõjudest teiste muutujatega (John et al. 1994: 128). Du Jardin'i (2012) arvates on muutujate ennustusvõime hindamine teiste muutujate mõju ning modelleerimise tehnikat teadmata ebamõistlik. See tähendab, et parim muutujate komplekt on mitmemõõtmeliste meetodite puhul valimispetsiifiline olles mõjutatud nii valimisse kaasatud vaatlustest, teistest finantsnäitajatest kui ka modelleerimise tehnikast.

Eelnevalt selgitas autor, milliste kriteeriumite alusel on eelnevas vastava ala kirjanduses uurimustöösse finantsuhtarve valitud. Järgnevalt tuleb ülevaade millised konkreetset finantssuhtarvud on varasemates kaubandussektorile keskendunud ettevõtete pankrotistumise prognoosimise mudelites oluliseks osutunud. Need finantssuhtarvud jagunevad järgnevasse autori poolt valitud kategooriatesse: lühiajalise maksevõime, efektiivsuse, rentaabluse, finantsstruktuuri ning varade struktuuri kajastavad näitajad.

Ettevõtte lühiajalise maksevõime ehk likviidsuse hindamisel kasutatakse mitmeid finantssuhtarve, millest enimkasutatud ettevõtete pankrotistumise prognoosimisel on lühiajalise võlgnevuse katterkodaja ehk käibevara suhe lühiajalistesse kohustustesse (*current ratio*), maksevõime kordaja (*quick ratio*) ning maksevalmiduskordaja (*cash ratio*). Lühiajalise maksevõime näitajad on olulised olukorras, mil omakapitali tootlus on madalam kui laenuintress, kuna sel juhul langeb ettevõtte suutlikkus kohustustusi täita. (Han et al. 2012:50). Tabelis 1.4 on toodud lühiajalise maksevõime näitajaid kaubandussektori mudelites kasutanud autorid ning meetodikad, mille abil mudel koostati. Kaubandussektoris on eelnevalt oluliseks osutunud eelpoolnimetatutest vaid käibevara suhe lühiajalistesse kohustustesse McGurri ja Devaney (1998) poolt.

Lühiajalise maksevõime näitajate prognoosivõime on kuni aasta pikkune, kuna käibevarade mahtu hoitakse ettevõttes ettenähtud piirides ning nende mahu muutmiseks

rahavoogude vähenemisel muutuvad vastavad finantsstruktuuri kajastavad näitajad, mille arvelt käibevarade mahtu muudetakse (Vuran 2009:59).

Tabel 1.4 Kaubandussektori uurimustes kasutatud lühiajalise maksevõime näitajad

Finantssuhtarv	Autor	Meetod
Käibevarade suhe lühiajalistesse kohustustesse	Mcgurr, Devaney 1998	Diskriminant analüüs
Äritegevuse rahavoo suhe kohustustesse	Lee, Choi 2013	Närvivõrkude meetod, diskriminant analüüs

Allikas: autori koostatud.

Lühiajalise maksevõime suhtarvud on lühiajalises plaanis ühed olulisemad indikaatorid ettevõtte seisundi kirjeldamiseks, kuna likviidsed vahendid on esimesed, mis raskustesse sattumisel puhvrina ära kasutatakse. Lisaks tekitab probleeme info jõudmine ettevõtte kesisest seisust võlausaldajateni, mille tõttu lõpetatakse krediidi andmine või lühendatakse maksetähtaegasid, mis omakorda vähendab likviidsete varade hulka. Likviidsete vahendite vähenemine kajastab Beaveri teeside järgi reservide vähenemist ning see saab juhtuda sissetulevate rahavoogude vähenemisel või väljaminevate rahavoogude suurenemisel, mille tõttu suureneb ettevõtte pankrotirisk.

Käibevarade suhe lühiajalistesse kohustusse on põhiline finantssuhtarv, mille abil hinnata kaubandusettevõtte suutlikkust lühiajaliste kohustuste tasumisel. Mida kõrgem on selle tase, seda väiksem võimalus on ettevõtte maksevõimetuse tekkimiseks. Mida suurem on ettevõtte kohustuste maht, seda keerulisem on ka pikaajaliste kohustuste tagasimaksete tegemine. (Mcgurr, Devaney 1998: 274)

Tabel 1.5 Kaubandussektori uurimustes kasutatud efektiivsuse suhtarvud

Finantssuhtarv	Autor ja aasta	Meetod
Müügitulu suhe koguvaradesse	Mcgurr, Devaney 1998	Diskriminant analüüs
Müügitulu suhe keskmisesse debitoorsesse võlgnevusse	Lukason 2006	Diskriminant analüüs, logit mudel
Müügitulu suhe keskmisesse käibekapitali mahtu	Lee, Choi 2013	Närvivõrkude meetod, diskriminant analüüs
12*kreditoorse võlgnevuse suhe müügitulusse	Shirata 1998	Diskriminant analüüs
Müüdud kaupade kulu suhe koguvaradesse	Bhargava et al. 1998	Logit mudel

Allikas: autori koostatud.

Efektiivsuse suhtarvude abil hinnatakse müügitulude suhet erinevatesse vara liikidesse ning nad iseloomustavad, kui hästi suudetakse müügitulu tekitada. Läbi nende tekib aimdus ettevõtte suutlikkusest konkurentsi tingimustes (Altman: 1968:595), seda võrrelduna sama haru ettevõtetega. Müügi suhe varadesse näitab juhtkonna võimekust kasutada ettevõttes olevaid ressursse. Näitajate puhul peaks selle kõrgem väärtus võrrelduna teiste ettevõtetega, tähendama väiksemat ebaõnnestumise riski ning näitama ettevõtte suuremat efektiivsust varade kasutamisel.

Suurima mõjuga näitaja McGurri ja Devaney mudelis oli müügitulu suhe koguvaradesse, mis nende arvates on üks enimkasutatud finantssuhtarv pankrotistumise prognoosimise alastes uurimustöodes. (Mcgurr, Devaney 1998: 273) Müügitulu rentaablus näitab ettevõtte põhitegevuse kasumlikkust ning erinevatesse vara tüüpidesse suhestatud müügitulu suhtarvud näitavad tegevuse efektiivsust. (Vuran 2009: 60)

Rentaabluse suhtarvud (vt. tabel 1.6) on ettevõtte erineva tasandi kasumlikkuse näitajate suhted erinevat tüüpi varadesse, peamiselt koguvaradesse ning ka müügitulusse. Nende puhul tähistab nende suurem tase ettevõtte suuremat edu ning seega nende kõrgem väärtus tähendab väiksemat pankrotistumise tõenäosust. (Han et al. 2012: 50)

Tabel 1.6 Kaubandussektori uurimustes kasutatud rentaabluse näitajad

Finantssuhtarv	Autor ja aasta	Meetod
Brutokasumi suhe koguvaradesse	Mcgurr, Devaney 1998	Diskriminant analüüs
Puhaskasumi suhe koguvaradesse	Bhargava et al. 1998	Logit mudel
	Mcgurr, Devaney 1998	Diskriminant analüüs
	Lukason 2006	Diskriminant analüüs , logit mudel
	Lee, Choi 2013	Närvivõrkude meetod, diskriminant analüüs
	Paal 2013	Logit mudel
Rahavoo suhe müügitulusse	Lukason 2006	Diskriminant analüüs, logit mudel
Äri kasumi suhe müügitulusse	Paal 2013	Logit mudel
Jaotamata kasumi suhe koguvaradesse	Shirata 1998	Diskriminant analüüs
	Lee, Choi 2013	Närvivõrkude meetod, diskriminant analüüs

Allikas: autori koostatud.

Vuran'i (2009: 60) arvates näitavad rentaabluse suhtarvud ettevõtte juhtimise efektiivsust. Nende abil saab hinnata ettevõtte kasumlikkust müügitulusse,

koguvaradesse või omakapitali mahtu. Omakapitali suhestatuna saab hinnata investeringute tasuvust. Rentaabluse näitajaid kasutatakse palju ettevõtete tulemuste hindamisel ning võrdlemisel ning seega on need esimesed näitajad, mida ettevõtete juhtkond võib moonutada, et näidata ettevõtte finantsseisu paremas valguses nii omanikele, nõukogule kui ka potentsiaalsetele investoritele.

Mcgurr ja Devaney (1998: 273) poolt püstitatud kaubandussektori pankrotistumise prognoosimise mudelis osutus oluliseks brutokasumi suhe koguvaradesse ning selle kõrgem väärtus tähendas väiksemat pankrotistumise tõenäosust. Kuigi varasemates antud valdkonna uuringutes pole see näitaja väga suurt olulisust omanud, siis kaubandussektori puhul on see oluline. (Mcgurr, Devaney 1998: 273) Autori arvates võimaldab see näitaja hinnata kaubandusettevõtete puhul nende hinnastamise efektiivsust. Kaubandussektori puhul on autori arvates kõik muud kulud oluliselt ühtlasemad, kuid näiteks tööstussektori ettevõtteid võrreldes pole see nii, kuna nende tootmistehnoloogiad ning nende kulu võivad sektorite üleselt üksteisest kordades erineda. Kaubandussektori kulude struktuur peaks autori hinnangul olema homogeensem ning seega on ka võimalik ettevõtteid omavahel paremini võrrelda.

Finantsstruktuuri näitajate (vt. tabel 1.7) põhjal on võimalik hinnata ettevõtte erinevate finantseerimisallikate osakaalu koguvaradest ja finantseerimisallikate dünaamikat kajastavaid näitajaid, mida kasutasid Mcgurr ja Devaney (1998) ning Shirata (1998) oma töödes. Du Jardin (2012: 22) hoiatab, et alla 4 aasta vanuste ettevõtete puhul võivad finantsstruktuuri näitajad olla sarnased ebaedukate ettevõtetega, kuna nende puhul pole see passiva struktuur veel jõudnud välja kujuneda.

Kaubandusettevõtetel, mille pikaajaliste kohustuste maht suurenes, oli ebaõnnestumise tõenäosus väiksem kui neil, millel see vähenes. See võib tunduda ebaloominguline, kuna pikaajaliste kohustuste mahu suurenemine tõstab maksejõuetuse riski, mis peaks sel juhul tõstma ka ebaõnnestumise tõenäosust. Pankrotistaval ettevõttel on väiksem tõenäosus saada täiendavat pikaajalist laenu, mis omakorda tähendab ka, et raskustes ettevõttel on sisuliselt võimatu suurendada pikaajaliste kohustuste mahtu. Seega

pikaajaliste kohustuste mahu positiivne dünaamika tähendab, et ettevõtte on krediidiuasutuste jaoks tulevikuväljavaate suhtes usaldusväärne ning sellega on põhjendatud pikaajaliste kohustuste mahu suurenemisel ettevõtte väiksem pankrotistumise tõenäosus. (Mcgurr, Devaney 1998: 273-274)

Tabel 1.7 Kaubandussektori uurimustes kasutatud finantsstruktuuri ning finantseerimisallikate dünaamika näitajad

Finantssuhtarv	Autor ja aasta	Meetod
Omakapitali suhe koguvaradesse	Paal 2013	Logit mudel
Pikaajaliste kohustuste suhe koguvaradesse	Mcgurr, Devaney 1998	Diskriminant analüüs
Pikaajaliste kohustuste muutus võrreldes eelneva perioodiga	Mcgurr, Devaney 1998	Diskriminant analüüs
Lühiajaliste kohustuste ja omakapitali summa muutus võrreldes eelneva perioodiga	Shirata 1998	Diskriminant analüüs

Allikas: autori koostatud

Varem on varade struktuuri näitajatest kaubandussektori ettevõtete pankrotistumise prognoosimise uurimustes kasutanud vaid antud töö autor. (vt. tabel 1.8). Seega võib eeldada, et nende kaasamine kaubandussektori ettevõtete pankrotistumise prognoosimise mudelitesse on pigem erand. Varade struktuuri näitajad kajastavad aktiva poole kirjade osakaalu bilansimahus.

Tabel 1.8 Kaubandussektori uurimustes kasutatud varade struktuuri näitajad

Finantssuhtarv	Autor ja aasta	Meetod
Raha ja pangakontode väärtuse suhe koguvaradesse	Paal 2013	Logit mudel, kaubandussektor

Allikas: autori koostatud.

Mudelites kasutatud ka suurust kajastavaid näitajaid, mida küll autor ei leidnud ühestki kaubandussektorile keskendunud uurimustööst.. Populaarsemad neist on omakapitali ja koguvarade maht, müügitulu ning töötajate arv. Nende näitajate muutumiskiirgus, võrrelduna finantssuhtarvudega, on väga lai ning seega kasutatakse neid mudelites teisendatuna, peamiselt logaritmituna.

Muutujate korduv kasutamine sama valkonna uurimustes ei tähenda, et need muutujad võiksid osutada oluliseks järgnevates uute andmetega läbi viidud uurimustöodes. Mudelid annavad parimaid tulemusi nende treenimiseks kasutatud andmetega ning mistahes kontrollvalimitega mudelite järelkontrollid annavad tavaliselt halvemaid tulemusi. Muutujate kirjeldusvõime mudelites sõltub teistest mudelisse kaasatud muutujatest ehk koosmõjudest. (du Jardin 2012: 21)

Seega muutujate valik uurimustöösse sõltub autori poolt seatud eesmärkidest ehk siis oluliseks peetavatest faktoritest ettevõtte tegevuse kirjeldamisel. Kui eesmärgiks on luua täiesti uus sektoripõhine pankrotistumise mudel, siis antud töö autori arvates on vajalik kaasata mudelisse võimalikult lai amplituud andmebaasi võimaluste piires erinevaid ettevõtte majandustegevust kajastavaid bilansi ja kasumiaruande näitajaid. Seejärel on võimalik kasutada statistilisi meetodeid, mille abil uurida muutujate kirjeldusvõimet ning ka muutujate gruppidevahelisi varieeruvuse erinevusi näiteks Brown& Forsythe testi abil. Järgmises peatükis on autori eesmärgiks koostada erinevaid finantssuhtarve kasutades autokaubanduse sektori ettevõtete pankrotistumise prognoosimise mudel, kasutades ühte klassikalist statistilist meetodit ning kahte masinõppimise meetodit.

2.EUROOPA AUTOKAUBANDUSETTEVÖTETE PANKROTI PROGNOOSIMISE MUDELITE KOOSTAMINE JA VÕRDLUS

2.1.Mudelite koostamiseks kasutatud andmed ja metoodika

Autori eesmärgiks on koostada Euroopa-ülesed NACE rev. 2 G45 sektori ehk automüügi ettevõtete pankroti prognoosimise mudelid, kasutades kolme erinevat meetodit, millest üks on klassikaline statistiline meetod ehk logistiline regressioon ja kaks neist masinõppimise meetodid ehk k-lähima naabri meetod ja juhusliku metsa meetod. Seejärel võrdleb autor saadud mudeleid omavahel nende erinevate klassifikatsiooni-täpsuse näitajate lõikes. Eesmärgiks on teada saada, milline mudel klassifitseerib pankrotistunud ettevõtteid kontrollvalimi põhjal kõige paremini. Lisaks toob autor antud peatükis välja mudelites olulised autokaubanduse sektori spetsiifilised finantssuhtarvud, millel on selle sektori ettevõtete ebaõnnestumise koha pealt kõige suurem ennustusvõime või mille kaasamata jätmine mudelitesse põhjustab kõige suuremat ebatäpsuse langust mudelite juures.

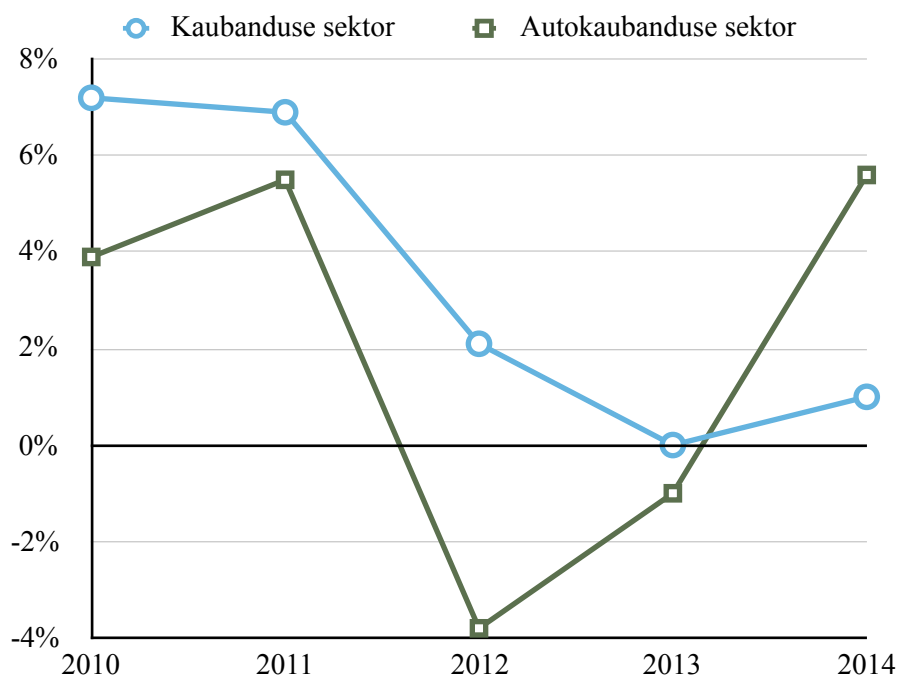
Enne andmete juurde siirdumist tutvustab autor automüügi sektorit ning selle osatähtsust Euroopa kaubandussektoris. Tabelist 2.1 nähtub, et automüügi sektori ettevõtted moodustavad 13% kõigist kaubandussektori ettevõtetest ning nende osakaal moodustab Euroopa ettevõtete koguarvust 3,6%. Tabelis toodud andmete kohaselt on automüügi ettevõtete koguarv ning osakaal kaubandusettevõtetest kasvavas trendis. Mootorsõidukite hoolduse ja parandusega tegelevaid ettevõtteid on üle poole antud sektoris tegutsevaist, mis teeb sellest kõige suuremat kaalu andva osa arvestades mudelite koostamist. Kõige väiksema osa moodustab mootorrataste müük, hooldus, parandus ning nendega seotud osade ja tarvikute müük.

Tabel 2.1 Euroopa Liidu kaubandussektoris tegutsevate ettevõtete arv tuhandetes

Euroopa Liit / aasta	2010	2011	2012	2013
Kaubandussektori ettevõtete arv tuhandetes	6 222	6 272	6 242	6 215
Mootorsõidukite ja mootorrataste hulgi- ja jaemüügi ning hoolduse ettevõtete arv tuhandetes	787	799	807	813
sh. mootorsõidukite müük	25,0%	25,4%	25,3%	25,4%
sh. mootorsõidukite hooldus ning parandus	56,2%	55,7%	56,0%	56,0%
sh. mootorsõidukite osade ning tarvikute müük	14,0%	14,2%	14,2%	14,2%
sh. mootorrataste müük, hooldus, parandus ning nendega seotud osade ja tarvikute müük	4,8%	4,6%	4,6%	4,4%

Allikas: Eurostat, autori koostatud

Autokaubanduse sektori ettevõtete hulgast on kõige vähem esindatud ettevõtete arvu poolest mootorrataste müügi, hoolduse, paranduse ning nendega seotud osade ja tarvikute müügiga tegelev alamsektor, seega suurema osa antud sektorist moodustavad mootorsõidukitesse puutuvad alamkategooriad. Alamsektorite osakaalud G45 sektorist on 2010...2013 perioodil olnud suhteliselt stabiilsed.



Joonis 2.1 Euroopa Liidu kaubanduse- ning autokaubanduse sektori müügikäibe muutus 2010...2014 (autori koostatud; Eurostat 2016 andmete põhjal).

Kaubandussektori käive moodustas 2013. aastal Euroopa liidu kõikide ettevõtete käibest 37,5%, mis tähendab, et autokaubanduse sektor samal aastal moodustas ~4% kogu ühenduse alal tegutsenud ettevõtete käibest (Eurostat 2016). Joonisel 2.1 on toodud

Euroopa Liidu kaubandussektori müügikäibe dünaamika aastatel 2010...2014 ning sellelt võib näha, et alamsektor reageeris vaadeldavatel aastatel kaubandussektori käibe muutusele oluliselt enam kui kaubandussektori käibe muutus. Mootorsõiduk on ühelt poolt tarbeesemeks, kuid antud dünaamika põhjal võib järeldada, et mootorsõiduk pole esmatarbekaup ning selle soetamist on võimalik negatiivse majandusliku tuleviku väljavaate ilmnemisel edasi lükata.

Kõik töös kasutatavad andmed pärinevad Bureau van Dijk'i Amadeusi andmebaasist. Sihiks on leida sektori ettevõtete kohta nii palju kasutatavaid pankrotistunud ettevõtete andmeid kui võimalik. Esialgsesse valimisse kuulusid ~57 000 pankrotistunud, likvideeritud või muul viisil tegevuse lõpetanud ettevõtte andmed. Ettevõtete staatuse muutmine (pankrot, likvideerimine pankroti tõttu) on toimunud vahemikus 2011...2016. aasta.

Tabel 2.2 Euroopa autokaubanduse ettevõtete pankrotistumise prognoosimise mudelite koostamiseks kasutatav valim riikide lõikes

Riik	Aktiivne	Pankrot	Kokku
Belgia	139	58	197
Hispaania	124	108	232
Horvaatia	38	22	60
Itaalia	345	403	748
Läti	12	27	39
Prantsusmaa	674	344	1018
Rumeenia	63	141	204
Serbia	13	6	19
Sloveenia	31	26	57
Soome	52	38	90
Tšehhi	41	12	53
Ukraina	18	41	59
Ungari	99	119	218
Vene Föderatsioon	22	22	44
Kokku	1671	1367	3038

Allikas: autori koostatud.

Edasisel analüüsil filtreeris autor välja kõigi nende ettevõtete andmed, mida analüüsis kasutada pole võimalik. Põhjusteks olid kas andmete puudulikkus või staatuse mittevastavus püstitatud eesmärgiga. Filtreeringu tulemusena jäi kasutatavasse andmebaasi 44 riigi ettevõtetest alles 14 riigi ettevõtte, mis on kajastatud tabelis 2.2.

Lisaks Euroopa Liidu ettevõtetele on valimis kajastatud EL-i mittekuuluvad riigid (Serbia, Ukraina ja Vene Föderatsioon), kuid need moodustavad valimist alla 5%. Kahjuks pole autori koostatud valimis suurriike: Saksamaa, Inglismaa. Valimi koostamise üheks tingimuseks oli, et ettevõtete koguvarade maht oleks vähemalt 200 tuhat eurot, et vältida tegevuseta või äärmiselt marginaalse äritegevusega ettevõtete sattumist valimisse. Põhjuseks on autori hinnangutel väiksemate ettevõtete finantsuhtarvude suur heterogeensus keskmisest, mis vähendab oluliselt mudelite võimalikke klassifikatsiooni määrasid. Siinkohal tuleb lisada, et antud andmebaas on pigem mikro ning väikeste ettevõtete kohta. Keskmise suurusega ettevõtteid Euroopa Komisjoni definitsiooni järgi on andmebaasis vaid 17 ning nendest pankrotistunud antud uurimustöös vaadeldaval perioodil on vaid 2.

Suure osa riikide puhul oli likvideerimise põhjus puudu. Selleks põhjuseks võis olla kas omanike soovil likvideerimine, ettevõtete ühinemise tõttu, nende jagamise tõttu või pankrotistumise tõttu. Kuna antud töös koostatakse ettevõtete pankrotistumise mudel, siis nendest likvideeritud ettevõtetest on võimalik kasutada ainult nende andmeid, mille puhul on likvideerimine toimunud pankrotistumise tõttu. Paljude riikide puhul oli välja toodud ka ettevõtete pankrot koos pankrotimenetluse alustamise ajaga.

Analüüsis kasutatud andmebaasis puuduvad mitmed Euroopa suur- ning väikeriigid. Autor toob siinkohal põhjused, miks pole andmebaasis kajastatud mitmeid suurriike. Märkusena tuleb lisada, et kõikide riikide puudumine andmebaasist on tingitud sarnasel põhjusel - polnud võimalik kindlaks teha pankrotistumist või selle täpset hetke. Saksamaa puhul on ettevõtete likvideerimise kuupäev, kuid pole võimalik kindlaks teha, millisel põhjusel likvideerimine toimus ning seega ei saa Saksamaad kasutada pankrotistumise mudelite loomisel. Ühendkuningriigi ning Hispaania ettevõtete puhul oli staatusena antud, et ettevõtted on likvideerimisel ning ka see pole piisav, et teada saada nende likvideerimise põhjust ning seega nimetatud riikide ettevõtted tuli analüüsist välja jätta.

Lõplikusse valimisse jäid 1367 pankrotistunud ettevõtte andmed. Seejärel leidis autor kõigi riikide pankrotistunud ettevõtetele juhusliku valiku alusel sarnase suurusjärgu aktiivseid ettevõtteid. Ettevõtete arv gruppides ei ole võrdsustatud kuna autor ei seadnud seda eesmärgiks omaette ning valik teostati juhuslikult. Aktiivsete ettevõtete puhul on teada, et nad eksisteerisid vähemalt 2014. aastaaruande esitamise ajal, järeltulevikult tegutsesid need ettevõtted veel vähemalt 2015. aasta suvel. Autor märgib ohuna, et ettevõtete puhul, mille aktiivne staatus ehk tegutsemine on eelduslikult teada ainult pool aastat peale viimast aruannet, võib suurendada mudeli II tüüpi vea esinemist. Need eeldusena aktiivsed ettevõtted, mis mudeli järgi klassifitseeritakse pankrotistunuks ehk tehakse II tüüpi viga, võivad ebaõnnestuda lähitulevikus.

Tabel 2.3 Valimi jaotus tegevusala põhiselt sh. nende ettevõtete staatus

Tegevusala	Aktiivne	Pankrotis	Kokku
G45.1 mootorsõidukite müük	660	646	1306
G45.2 mootorsõidukite hooldus ning parandus	609	357	966
G45.3 mootorsõidukite osade ning tarvikute müük	317	270	587
G45.4 mootorrataste müük, hooldus, parandus ning nendega seotud osade ja tarvikute müük	85	94	179
Kokku	1671	1367	3038

Allikas: Autori koostatud.

Valitud sektor jaguneb mitmeks alamsektoriks, mis esialgse analüüsi järgi tunduvad olevat suhteliselt erinevad, kuna sektor jaguneb mootorsõidukite hulgi-, jaemüügi ja hooldusega tegelevateks ettevõteteks, mis autori hinnangul ettevõtte bilansi struktuuride mõistes on üksteisest erinevad. Tabelis 2.3 on toodud sektori täpsem jaotumine alamsektoriteks ning tegutsevate ja pankrotistunud ettevõtete jaotus. Antud valimi puhul on enim pankrotistunud ettevõtteid esindatud mootorsõidukite müügiga tegelevas alamsektoris, mida on 47% kõigist pankrotistunud ettevõtetes. Tabelis 2.1 moodustasid mootorsõidukite müügiga tegelevad ettevõtted 25% kõigist autokaubanduse sektori ettevõtetest. Seega kuulub käesolevasse valimisse proportsionaalselt rohkem mootorsõidukite müügiga tegutsevaid ettevõtteid kui seda on autokaubanduse sektoris. Aktiivsete ning pankrotistunud ettevõtete vahekord kõigi tegevusalade lõikes jääb autori poolt eeldatud piiridesse, mille puhul gruppide suurus ei erineks üksteisest rohkem

kui poole võrra. See võimaldab järgnevas töös koostada kõigi gruppide põhjal eraldi logit mudel.

Tabel 2.4 Autori poolt töösse valitud muutujad, nende kategooriad ning arvutusvalemid

Kategooria	Suhtarv	Arvutusvalem
Maksevõime	makse1	Käibevarad / lühiajalised kohustused
	makse2	Raha / lühiajalised kohustused
Likviidsuse	liq1	(Käibevarad – varud) / lühiajalised kohustused
	liq2	Raha / koguvarad
	liq3	Raha / (lühiajalised kohustused + pikaajalised kohustused)
	liq4	EBITDA / müügitulu
Rentaablus	rent1	(Aruandeaasta kasum - muud äritulud) / äritulud kokku
	rent2	Äritulud kokku / koguvarad
	rent3	(Aruandeaasta kasum - muud äritulud) / koguvarad
	rent4	Aruandeaasta kasum / omakapital
	rent5	EBITDA / koguvarad
	rent6	EBITDA / põhivarad
	rent7	EBIT + finantstulu / koguvarad
	rent8	EBIT / koguvarad
	rent9	Aruandeaasta kasum / koguvarad
Finantsstruktuur	fin1	Pikaajalised kohustused / koguvarad
	fin2	Kohustused / koguvarad
	fin3	Lühiajalised kohustused / koguvarad
	fin4	Omakapital / koguvarad
	fin5	Põhivarad / omakapital
	fin6	Põhivarad / (omakapital + pikaajalised kohustused)
	fin7	Pikaajalised kohustused / omakapital
	fin8	Kohustused / omakapital
Kapitalstruktuur	kapstr1	Varud / koguvarad
	kapstr2	Käibevarad / koguvarad
	kapstr3	(Raha + varud) / koguvarad
	kapstr4	(Varud + nõuded ja ettemaksed) / koguvarad
Efektiivsus	ef1	Äritulud kokku / omakapital
	ef2	(Aruandeaasta kasum + amortisatsioon) / koguvarad
	ef3	EBIT / müügitulu
	ef4	Lisandväärtus / müügitulu
	ef5	EBITDA / kohustused
	ef6	EBITDA / intressi kulud
	ef7	Müügitulu / käibevarad
	ef8	(Aruandeaasta kasum + amortisatsioon) / müügitulu
Suurus	ta_1	Koguvarad

Allikas: autori koostatud.

Muutujate valikul on autor seadnud esimeseks kriteeriumiks, et oleks kaetud võimalikult suur amplituud erinevaid ettevõtte tegevuse aspekte kajastavaid suhtarve.

Autor soovib kajastada järgnevaid finantssuhtarvude kategooriaid, mis on hübriid Vurani (2009) ning Jardini (2012) väljapakututest. Nendeks on lühiajalist maksevõimet, rentaablust, finantsstruktuuri, kapitalistruktuuri ning efektiivsust kajastavad näitajad, mis on toodud tabelis 2.4.

Tabelis 2.4 on kajastatud kõik töösse kaasatud muutujad koos nende nime, seletuse ning arvutusvalemiga. Muutujate nimi kajastab ka nende muutujate liiki. Maksevõime näitajad on kodeeritud “makse1...2” ning “liq1...4”. Rentaablust kajastavad näitajad on kodeeritud “rent1...9” . Finantsstruktuuri kajastavad näitajad “fin1...8” ning kiiresti käibivate varade finantseerimise struktuuri kajastavad näitajad “kapstr1...4”. Efektiivsust kajastavad näitajad on kodeeritud “ef1...ef8”. Arvutusvalemite kasutatav sufiks “_n” tähistab mitu aastat (1...2) varasemast ajast enne pankrotistumist pärineb aruanne ebaõnnestunud ettevõtete puhul. Kõikide muutujate puhul on autor arvutanud kaks perioodi ehk t-1 ja t-2, mis kujutavad vastavalt eelneva ning üleelneva majandusaasta aruandeid.

Antud töös tehtavast analüüsist jäetakse välja ettevõtte suurust kajastavad näitajad peale ettevõtte koguvarade mahu. Suurust kajastavate muutujate kaasamise miinuseks on, et võrreldes finantssuhtarvudega on nende võimalik muutumiskiirkond oluliselt laiem ning need võivad seetõttu mudelid niiöelda kiiva ajada. Lahenduseks on sel juhul muutujate teisendamine - näiteks logaritmimine. Ohuks neid kaasates on suur multikollineaarsuse oht seotuna teiste suurust kajastavate muutujatega. Ettevõtte suurust kajastavate näitajate mudelites kasutamise võimalikud negatiivsed küljed on autori enda varasemate katsetustele põhinev kogemus bakalaureusetöös mudelisse treenimisel kasutatud muutujate põhjal. Mudelite treenimisel proovitakse ainult ühte suurust kajastavat muutujat, milleks on koguvarade maht.

Andmaks paremat ettekujutust andmebaasist ning seal kajastatud muutujatest ja nende keskmiste väärtuste erinemisest gruppide lõikes, toob autor tabelis 2.5 kirjeldavad statistikud kõigi lõplikes mudelites kasutatud muutujate kohta. Kõigi muutujate puhul on autor tasandanud ekstreemumväärtused, võrdsustades alumise ja ülemise protsentiili

väärtused vastavalt protsentiili alumise protsentiili suurima väärtusega ning ülemise protsentiili kõige väiksema väärtusega. Kui finantsuhtarvudel esines puuduvaid väärtuseid alla 1%, siis asendas autor need väärtused vastavate staatuse gruppide (aktiivsed/pankrotistunud) keskväärtustega.

Tabel 2.5 Lõplikes mudelites kasutatud muutujate kirjeldavate statistikute võrdlus aktiivsete ja pankrotistunud ettevõtete vahel,

Muutuja	Aktiivsed ettevõtted (1671)				Pankrotistunud ettevõtted (1367)				Brown& Forsythe p-väärtus
	Keskmine	Standard-hälve	miinimum	maksimum	Keskmine	Standard-hälve	miinimum	maksimum	
ta_1	2810,2	15547,3	200,2	530125,0	2074,7	4478,6	200,0	60617,1	0,08
makse1_1	2,55	5,04	0,00	48,19	1,26	3,53	0,00	48,19	0,00
makse2_1	0,58	1,53	0,00	11,92	0,10	0,63	0,00	11,92	0,00
liq1_1	1,51	3,16	0,00	27,83	0,70	1,99	0,00	27,83	0,00
liq2_1	0,12	0,16	0,00	1,00	0,04	0,09	0,00	0,97	0,00
liq3_1	0,43	1,15	0,00	9,08	0,05	0,20	0,00	4,24	0,00
rent1_1	0,02	0,09	-0,74	0,42	-0,20	0,37	-1,93	0,42	0,00
rent2_1	0,04	0,10	-0,64	0,51	-0,17	0,35	-1,85	0,51	0,00
rent3_1	0,02	0,09	-0,74	0,60	-0,22	0,53	-11,00	0,60	0,00
rent5_1	0,07	0,10	-0,61	0,61	-0,12	0,36	-4,08	0,61	0,00
rent7_1	0,03	0,11	-0,74	0,73	-0,22	0,53	-11,00	0,73	0,00
rent8_1	0,04	0,10	-0,64	0,61	-0,19	0,50	-10,65	0,74	0,00
rent9_1	0,03	0,09	-0,74	0,54	-0,25	0,59	-10,94	0,64	0,00
fin1_1	0,15	0,21	0,00	1,88	0,29	0,39	0,00	4,13	0,00
fin2_1	0,66	0,30	0,00	3,20	1,33	1,14	0,00	12,55	0,00
fin3_1	0,51	0,29	0,00	3,18	1,05	1,08	0,00	12,35	0,00
fin4_1	0,34	0,30	-2,20	1,00	-0,33	1,14	-11,55	1,00	0,00
fin5_1	1,46	4,74	-19,77	31,33	0,86	6,84	-19,77	31,33	0,00
fin6_1	0,79	2,37	-14,56	16,38	0,53	3,90	-14,56	16,38	0,00
fin7_1	1,14	4,27	-14,05	29,11	0,78	6,12	-14,05	29,11	0,00
fin8_1	4,96	17,95	-67,90	119,75	3,80	25,95	-67,90	119,75	0,00
kapstr1_1	0,31	0,25	0,00	1,00	0,29	0,26	0,00	1,00	0,13
kapstr2_1	0,71	0,26	0,00	1,00	0,68	0,29	0,00	1,00	0,00
kapstr3_1	0,44	0,26	0,00	1,00	0,33	0,27	0,00	1,00	0,11
kapstr4_1	0,26	0,26	-1,08	0,98	0,14	0,37	-1,94	1,00	0,00
ef1_1	11,05	31,72	-98,94	274,20	9,94	47,82	-98,94	274,20	0,00
ef2_1	0,05	0,09	-0,65	0,55	-0,16	0,43	-5,02	0,55	0,00
ef5_1	0,18	0,35	-1,08	2,50	-0,05	0,22	-1,08	2,50	0,00
ef7_1	1,85	1,52	0,00	18,51	1,70	3,22	0,00	75,89	0,00
ef8_1	2,78	2,82	0,00	53,38	2,71	4,73	0,00	65,83	0,00

Allikas: Bureau van Dijk'i Amadeus'i andmebaas, autori arvutused.

Pankrotistunud ja aktiivsete ettevõtete kõrvutamisel selgub, et rentaabluse näitajate keskmised pankrotistunud ettevõtete puhul on kõik negatiivsed väärtused ehk keskmine

pankrotistunud ettevõtte tegutseb kahjumlikult, mis on loogiline ning aktiivsete ettevõtete puhul kirjeldavad kaubandussektorile omast mitte väga kõrget rentaablust jäädes vahemikku 0,02...0,07. Brown ja Forsythe testi tulemusena saab öelda, et vaid 3 muutuja puhul on pankrotistunud ning aktiivsete ettevõtete vastavate gruppide hajuvus ühesugune ning need on tabelis 2.5 rasvases kirjas toodud olulisustõenäosustena $> 0,05$. Nendeks muutujateks on ta_1 ehk koguvarad perioodil $t-1$, $kapstr1_1$ ehk varud / koguvarad perioodil $t-1$ ja $kapstr3_1$ ehk $(\text{raha} + \text{varud}) / \text{koguvarad}$ perioodil $t-1$.

Järgnevates alapeatükkides kasutatakse G45 üleeuroopaliste pankrotimudelite koostamisel kolme viisi: logit mudel, k-lähima naabri meetod ning juhuslik mets (*random forest regression*). Mudelite koostamisel kasutatakse tarkvarapakette Stata ning R. Kõigi meetodite puhul kasutatakse mudelite loomisel treeningvalimit ning testitakse seejärel mudeli headust kontrollvalimi abil, mis annab võimaluse koostatud mudelite võrdlemiseks. Kõigi mudelite kohta toob autor välja üldise klassifitseerimismäära, tundlikkuse ja spetsiifilisuse. Kuna tegemist on andmeanalüüsi mõistes küllaltki erinevate statistiliste meetoditega, siis pole võimalik mudelite võrdluseks kasutada spetsiifilisi mudelite hindamise kriteeriume nagu näiteks logit/probit mudeli jaoks Akaike informatsioonikriteerium. Seega mudelite omavaheliseks võrdluseks tuleb kasutada erinevate kriteeriumide alusel hindamist. Lisaks on võimalik jagada valim näiteks riigi- või tegevusala põhiselt ning läbi selle on võimalik hinnata mudeli stabiilsust.

2.2.Euroopa autokaubanduse ettevõtete pankrotistumise prognoosimine logistilise regressiooni meetodil

Järgnevas analüüsis on autor jaganud valimi juhuslikkuse alusel kaheks: treening- ning kontrollvalimiks, mis on täpselt sama mahuga. Esimesse kuulub 1519 ettevõtet, millest 825 on aktiivsed ning 694 on pankrotistunud. Kontrollvalimis on 846 aktiivse ning 673 pankrotistunud ettevõtte andmed, mis peaks olema autori hinnangul piisav, võimaldamaks kontrollvalimi põhjal teha üldistava järelduse mudelite headusele. Järgnevalt koostatakse mudelid, mis baseeruvad treeningvalimil ning kontrolliks

tuuakse ka klassifitseerimistäpsused kontrollvalimi puhul. Selles alapeatükis koostatakse Euroopa mootorsõidukite ja mootorrataste hulgi- ja jaemüügi ning hoolduse ettevõtete sektori kohta logit mudel.

Esimeseks etapiks logit mudeli koostamisel on multikollineaarsuse statistikute arvutamine valimile ning sammhaaval multikollineaarsust põhjustavate muutujate eemaldamine, kuni ühegi muutuja VIF ei ületaks väärtust 4 (O'Brien 2007: 684). Tabelis 2.6 on multikollineaarsuse statistikute tabel treeningvalimis olevatele muutujatele, millest on eemaldatud kõik multikollineaarsust põhjustanud muutujad vastavalt suurima VIF väärtusega muutuja eemaldamisega igal sammul, ning komplekt, mille põhjal algab esialgse logit mudeli hindamine ning järk-järgult muutujate eemaldamine selle mudeli klassifikatsiooniparameetrite parendamiseks.

Tabel 2.6 VIF väärtused muutujatele multikollineaarsuse hindamiseks, tabelis toodud muutujad, mille korral $VIF < 4$. Väärtused arvutatud "treeningvalimi" põhjal

Muutuja	VIF	Tolerants	R ²
liq1_1	1,78	0,56	0,44
liq2_1	2,00	0,50	0,50
liq3_1	3,02	0,33	0,67
rent1_1	3,56	0,28	0,72
rent2_1	1,77	0,57	0,43
fin1_1	1,57	0,64	0,36
fin4_1	2,44	0,41	0,59
fin5_1	2,89	0,35	0,65
fin6_1	1,46	0,68	0,32
fin7_1	2,33	0,43	0,57
kapstr1_1	1,86	0,54	0,46
kapstr2_1	2,26	0,44	0,56
kapstr4_1	1,6	0,62	0,38
ef1_1	1,77	0,56	0,44
ef2_1	3,19	0,31	0,69
ef5_1	2,00	0,50	0,50
ef8_1	1,74	0,58	0,42

Allikas: autori koostatud.

Eesmärgiks on saada parim logit mudel, milles oleks kaasatud <10 finantsuhtarvu, mudeli kõigi muutujate olulisustõenäosused oleksid <0,05, mudel oleks statistiliselt oluline ning kõigi kaasatud muutujate ees olevad kordajad oleks majandusteoreetiliste

tõekspidamiste ning formaalloogilise analüüsi põhjal õiged. Kui peaks juhtuma, et mõne ees olevate märkide suhtes tekib kahtlus, siis autor eemaldab antud muutujad mudelist.

Logit mudeli koostamisel võttis autor eesmärgiks, et maksimeerida lõpliku mudeli üldist klassifikatsioonimäära. Teise eesmärgina võib nimetada, et lõplikus mudelis oleks alla 10 muutuja. Mudelite hindamine toimus järkjärgult mudelist muutujaid eemaldades. Esimeseks eemaldamise kriteeriumiks oli muutuja olulisustõenäosus ning kõige suurema olulisustõenäosusega muutuja eemaldati igal sammul. Sellest hetkest, mil kõik muutujad olid olulisustõenäosusega $< 0,01$ hakkas autor muutujaid eemaldama multikollineaarsuse testi statistiku VIF järgi, mille korral suurima VIF väärtusega muutuja eemaldati mudelist. Parimas logit mudelis, mille parameetrite hinnangud on toodud tabelis 2.7, oli keskmine VIF statistiku väärtus 1.19 ning maksimaalne VIF statistiku väärtus muutujatest muutujal “rent1_1” 1.31, mida võib pidada näitajaks, mis peaks vältima multikollineaarsuse ohu mudelites, milles neid muutujaid kasutatakse.

Tabel 2.7 G45 sektori logit mudeli parameetrite hinnangud, marginaalsed efektid ning olulisustõenäosused treeningvalimi põhjal

Muutuja	Parameetrite hinnangud	Marginaalsed efektid	$P > z $
liq2_1	-5,272	0,005	0,00
rent1_1	-5,525	0,004	0,00
fin4_1	-2,873	0,057	0,00
kapstr4_1	-1,029	0,357	0,00
vabaliige	0,514	1,671	0,00

Allikas: autori koostatud.

Andmebaasis on pankrotistunud ettevõtete puhul sõltuvaks muutuja väärtuseks määratud pankrotistunud ettevõtete puhul 1 ning aktiivsetel ettevõtetel 0. Kõigi mudelisse kaasatud muutujate (tabel 2.7) parameetrid on majandusteoreetiliste eeldustega kooskõlas, mis näitab veelkord, et mudelis ei tohiks esineda multikollineaarsust, mille esinemist VIF statistikute väärtusele tuginedes püüti vältida. Mudeli marginaalsetest efektidest võib järeldada, et muutuja “kapstr4_1” ehk “(Varud + nõuded ja ettemaksed) / koguvarad” ühe ühikuline muutumine mõjutab enim pankrotistumise tõenäosuse muutumist 35,7% võrra. Seega võib öelda, et mudel on kõige tundlikum muutuja “kapstr4_1” muutumisele. Teiseks muutujaks, mis mudeli

tundlikkusse enim panustab, on omakapitali suhe koguvaradesse, mille üheühikuline suurenemine/vähenedmine suurendab/vähendab pankrotistumise tõenäosust 5,7% võrra.

Parima logit mudeli üldkuju on defineeritud järgnevalt (kus K_1 on skoor

pankrotistumise tõenäosuse avaldises $p = \frac{1}{1 + e^{-K_1}}$):

$$(1.1) \quad K_1 = -5,272 \frac{\text{raha}}{\text{koguvarad}} - 5,525 \frac{\text{aruandeaasta kasum} - \text{muud äritulud}}{\text{äritulud kokku}} - 2,873 \frac{\text{omakapital}}{\text{koguvarad}} - 1,029 \frac{\text{varud} + \text{nõuded ja ettemaksed}}{\text{koguvarad}} + 0,514$$

Mudeli stabiilsuse kontrolliks hindas autor tabelis 2.7 toodud muutujatega logit mudeli kontrollvalimi põhjal ning muutujate kordajad olid samade märkidega ehk kõigi kordajate ees oli miinusmärk, välja arvatud vabaliikmel, ning parameetrite hinnangud olid sarnases suurjajärgus ehk ei erinenud rohkem kui 15% võrra treeningvalimi puhul koostatud mudelist. See annab alust oletada, et koostatud mudel on suhteliselt stabiilne. Kõikide muutujate ees olev miinusmärk on kooskõlas ka majandusteoreetiliste arusaamadega, mille kohaselt nende finantssuhtarvude madalam tase tähendab kesisemat finantsseisundit.

Mudeli koostamisel testis autor lisaks eelneva (t-1) perioodi andmetele ka üleelamise (t-2) perioodi andmeid ehk valis kõigi tabelis 2.7 toodud muutujate puhul t-2 perioodist pärit andmed (tabelis 2.7 on t-1 perioodist andmed), kuid need andsid halvemaid tulemusi, mis on läbiv ka vastava teema varasemate uuringute puhul - parim ennustusvõime on t-1 perioodi finantsnäitajatel. Tulemused t-2 perioodist (1a 6kuud... 2a 5kuud) andmetega oli “treeningvalimi” puhul klassifikatsioonimääraks 75,38% ning “kontrollvalimi” puhul 74.65%, mis on mõnevõrra madalamad tulemused kui t-1 perioodi (6kuud...1a 5kuud) andmetega.

Logit mudeli klassifikatsioonitäpsus selle alamsektorite erineva struktuuri mõttes keerulise sektori puhul oli autorile ootuspärane (vt. tabel 2.8). Antud sektori puhul raskendab klassikalise statistilise mudeli koostamist alamsektorite erinevus üksteisest

bilansi- või kasumiaruannete struktuuri põhjal. Kui G45.1 (vt. tabel 2.3) sektor tegeleb mootorsõidukite müügiga, siis võib eeldada, et selles sektoris on rentaablus väikseim võrrelduna teiste alamsektoritega. G45.2 ehk hoolduse ning paranduse puhul on tööjõukulude suur osakaal käibest see, mis eristab seda teistest alamsektoritest. Varuosade ning tarvikute müügi puhul võib eeldada, et laovarude osakaal võrreldes teiste alamsektoritega on suurem. G45.4 sektori puhul võib eeldada, et sektori sarnaneb enim ülemsektorile G45, milles pole vastavad alamtegevused kategooriatesse jaotatud.

Tabel 2.8 Logit mudeli klassifitseerimistäpsused alamsektorite lõikes, kontrollvalimi põhjal, lõikepunkt (0,5)

Sektor:	G45.1		G45.2		G45.3		G45.4		G45	
Tegelik\mudel	0	1	0	1	0	1	0	1	0	1
0	302	47	240	44	124	23	38	7	704	121
1	86	234	67	118	46	100	7	36	206	488
Spetsiifilisus	86,53%		84,51%		84,35%		84,44%		85,33%	
Tundlikkus	73,13%		63,78%		68,49%		83,72%		70,32%	
Üldklassifitseerimine	80,12%		76,33%		76,45%		84,09%		78,47%	

Allikas: Autori koostatud.

Spetsiifilisus tähendab mudeli poolt õigesti aktiivseks klassifitseeritud ettevõtete suhet kõigist mudelisse kaasatud aktiivsetest ettevõtetest. Tundlikkus on mudeli poolt õigesti pankrotistunuks klassifitseeritud ettevõtete osakaal kõigist mudelisse kaasatud pankrotistunud ettevõtetest. Üldine klassifikatsioonimäär on kõigi mudeli poolt õigesti klassifitseeritud ettevõtete osakaal mudelisse kaasatud ettevõtetest ning see on antud juhul 78,47%, mis on logit mudeli kohta suhteliselt ootuspärane tulemus. Arvestada tuleks siinkohal, et üldvalimist on eemaldatud ettevõtted, mille puhul olid andmed puudu või käive oli alla 200 tuhande euro, et välistada liialt väikeste ettevõtete valimisse kaasamist. Mudelisse on kaasatud kõik pankrotistunud ettevõtted, mida antud sektoris kasutada oli võimalik, siis aktiivsete ettevõtete valik on tehtud juhusliku valiku põhimõttel ning mudeli klassifikatsioonimäära kunstliku tõstmise eesmärgil pole neid sobitatud ebaõnnestunud ettevõtetega. Sellega soovis autor saavutada reaalsele olukorrale võimalikult sarnaseid tingimusi ning sellest ka logit mudeli ootuspärane klassifikatsioonimäär.

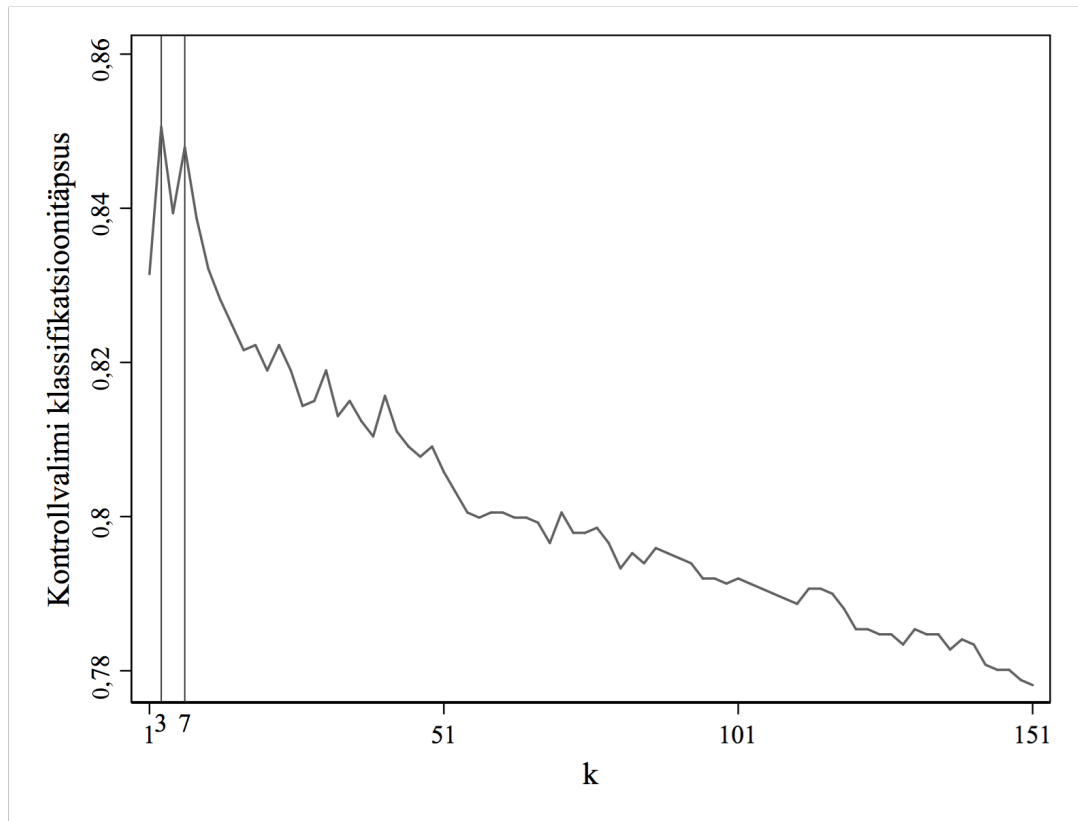
2.3.Euroopa autokaubanduse ettevõtete pankrotistumiste modelleerimine masinõppimise meetodeid rakendades

Kui logit mudeli puhul olid olulised muutujad mitte vanemad kui t-1 perioodi majandusaasta aruanne, siis masinõppimise meetodite puhul on autor kasutatud ka vanemaid andmeid. Andmed on eristatud üksteisest sufiksiga “*_1” (vt tabel 2.7), mis märgivad aruande esitamise ning pankrotistumise ajalist kaugust kuuest kuust kuni ühe aasta ja viie kuuni või sufiksiga “*_2”, mille puhul ajaline kaugus pankroti hetkest on enam kui üks aasta ja kuus kuud ning kuni kaks aastat ja viis kuud. Järgnevate mudelite treenimise juures kaasatakse kahest erinevast perioodist enne pankrotistumist pärit andmeid.

Seejärel viis autor läbi k-lähima naabri meetodil analüüsi treeningvalimi peal. Misjärel testis saadud tulemusi kontrollvalimil. K-lähima naabri meetod on nn “musta kasti” meetod ning selle mudeli kohta pole head graafilist väljundit. Kuna antud juhul kasutab autor mudelit dihhotoomse tunnuse jaoks klassifikatsioonimudeli koostamiseks, siis k väärtusteks on mõistlik valida paaritud arvud, et ei tekiks võimalust valikul võimalust viigiolukorra tekkimiseks. Viigiolukord tähendab, et mudel klassifitseerib vaatluse kahte gruppi kuuluvaks ning otsustamiseks kasutatakse sel juhul juhuslikku valikut (Muru 2015: 13).

Antud mudeli treenimiseks kasutas autor 58 finantssuhtarvu kahest erinevast perioodist (t-1 ning t-2 perioodi finantssuhtarvud). Katsetus tehti ka koguvarade kasutamisel antud mudelis, kuid kuna suhtarvude väärtused jäävad enamasti vahemikku 0...10, siis koguvarasid kirjeldav muutuja viis mudeli äärmiselt kiiva selle muutuja suurte absoluutväärtuste tõttu, mis tingis selle, et teiste muutujate panus vaatluste kirjeldamisel kahanes olematuks. See omakorda kahandas oluliselt õigesti klassifitseerimise määra ning seega pole ettevõtte suurust kirjeldav muutuja lõplikus mudelis kasutusel. Parima mudeli leidmine andme-komplektile toimub parima k väärtuse leidmise teel testvalimile. Selle meetodi puhul oli k väärtuseks läbiviidud analüüsis 3, mis tähendab, et igale punktile 3 kõige lähemat punkti n -dimensioonilises ruumis klassifitseeritakse temaga samasse klassi kuuluvaks. Dimensioonide hulk “ n ” on võrdne muutujate arvuga

ning vaatluste vahelise distantsti mõõtmiseks kasutati siinjuures Eukleidese distantsti. Autor kontrollis parima k -väärtuse leidmiseks väärtusi $1 \dots 150$ ning leidis, et parimaid tulemusi antud valimi põhjal andsid k väärtused < 9 , mille puhul ka nii treening kui ka kontrollvalimi õigesti klassifitseerimise määr olid suhteliselt sarnastes suurusjärgudes.



Joonis 2.2 K-lähima naabri meetodi kontrollvalimi klassifitseerimistäpsuse sõltuvus k väärtusest (autori koostatud).

Parima tulemuse andis treeningvalimi korral k väärtus 1, mille puhul oli klassifikatsioonitäpsus treeningvalimi puhul 100%, kuid autori hinnangul oli sellel puhul tegu küllastunud mudeliga (*overfitting*), mis tähendab, et mudel järgis äärmiselt täpselt treeningvalimit, kuid ei suutnud kontrollvalimi puhul olla piisava üldistusvõimega. Autor võttis eesmärgiks maksimeerida kontrollvalimi klassifikatsioonitäpsust ning parima tulemuse andis k väärtus 3, mille puhul kontrollvalimi õigesti klassifitseerimise määr oli parim (vt joonis 2.2).

Autor koostas läbi k -lähima naabri meetodi puhul mudelid k väärtustel $1 \dots 151$, järgides printsiipi, et kõik k väärtused oleks paaritud arvud. Jooniselt 2.2 nähtub, et mudeli jaoks

parim klassifitseerimistäpsus oli k väärtuse 3 juures ning järgnes k väärtus 7, mille mõlema puhul oli üldine klassifitseerimistäpsus kontrollvalimi korral üle 84%. K väärtuse suurenedes klassifitseerimistäpsus kontrollvalimi jaoks vähenes suhteliselt ühtlaselt kuni k väärtuseni 151, mille puhul klassifitseerimistäpsus on ~78%. Autor ei peatu antud uurimustöös k-lähima naabri meetodi klassifitseerimistäpsuste hindamisele suuremate k väärtuste korral, kuna see pole antud uurimistöö eesmärgiks ning ei anna paremaid tulemusi.

Tabel 2.9 K-lähima naabri meetodi klassifitseerimistäpsused sektorite lõikes, kontrollvalimi põhjal

Sektor:	G45.1		G45.2		G45.3		G45.4		G45	
Tegelik\mudel	0	1	0	1	0	1	0	1	0	1
0	329	20	265	19	137	10	44	1	775	50
1	80	240	51	134	40	106	6	37	177	517
Spetsiifilisus	94,27%		93,31%		93,20%		97,78%		93,94%	
Tundlikkus	75,00%		72,43%		72,60%		86,05%		74,50%	
Üldklassifitseerimine	85,05%		85,07%		82,94%		92,05%		85,06%	

Allikas: Autori koostatud.

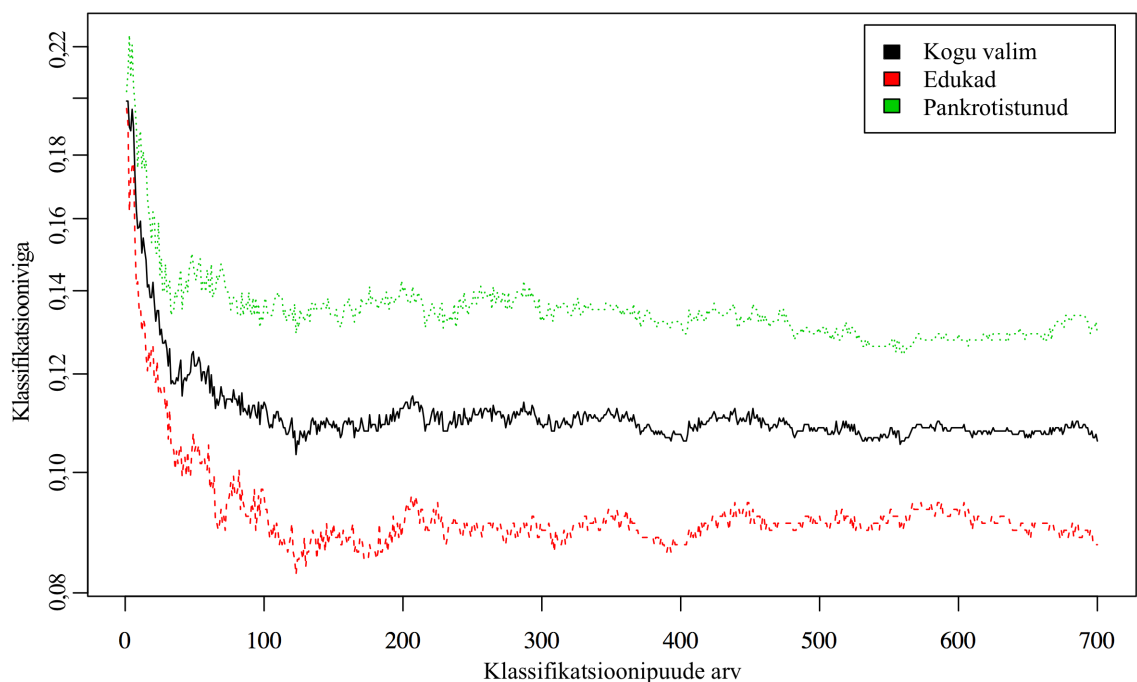
Klassifikatsioonitäpsused sektorite lõikes kontrollvalimi põhjal on toodud tabelis 2.9, milles on võimalik näha mudeli klassifikatsioonitabeleid alamsektoriti. K-lähima naabri meetodi puhul on alamsektorite klassifitseerimistäpsused suhteliselt lähedased kogu mudeli klassifikatsioonitäpsusele, kui välja arvata sektor G45.3, kuid autori arvates pole selle kohta paslik teha suuremaid järeldusi, kuna selle sektori maht on suhteliselt väike ning pigem on antud sektori kõrge klassifitseerimismäär juhuslik.

Juhusliku metsa meetod (*random forest*) on üks masinõppimise meetoditest. Autor valis meetodi, kuna ka selle puhul polnud andmestikule seatud eeldusi ning pole probleemiks ka suur muutujate arv. Kui k-lähima naabri meetodi puhul oli üheks võimalikuks mudeli parameetriks k väärtus ning vastavalt sellele mudeli parima tingimuseks seatud klassifikatsioonimäära poole püüdlemine, siis juhusliku metsa puhul on sarnaseks, kuid mitte samasuguseks parameetriks juhuslikus metsas kasvatatavate puude arv. Juhusliku metsa meetodi puhul koostab mudel etteseadud arvu klassifikatsioonipuid, millest igäühe koostamise puhul kasutatakse ~2/3 treeningvalimi mahust. Vastav osa valitakse

juhuslikult ning mudeli reprodutseerimiseks saab statistikatarkvarades kasutada nii juhuslikkuse genereerimise seemet (*random seed*), mis antud töös kasutatud mudelite puhul on “12345”, ning mida kasutades on võimalik antud töös olev mudel taasluua.

Siinkohal tuleb lisada, et kokku oli mudeli treenimisel kasutuses 58 suhtarvu kahest perioodist ning lisaks ettevõtte t-1 perioodi koguvarade mahtu kirjeldav muutuja. Viimase kasutamisel meetodiga, milles arvutatakse väärtus, mille alusel toimub jaotus oksadele, ei ole muutuja keskmise väärtuse erinevus teistest mudelis kasutatavatest muutujatest oluline.

Meetodi puhul on võimalik näha, kuidas muutub keskmine ennustusviga (*out of bag error*) iga järgneva puu genereerimisel. Peale 50 puu genereerimist hakkab keskmine ennustusviga koonduma 0,11 juurde mida võib näha joonisel 2.3. Sellel on näidatud ka keskmise ennustusvea muutumist gruppide lõikes, ning antud juhul ennustusvea paranemist ning stabiliseerumist vastavalt klassifitseerimispuude hulga kasvamisega.



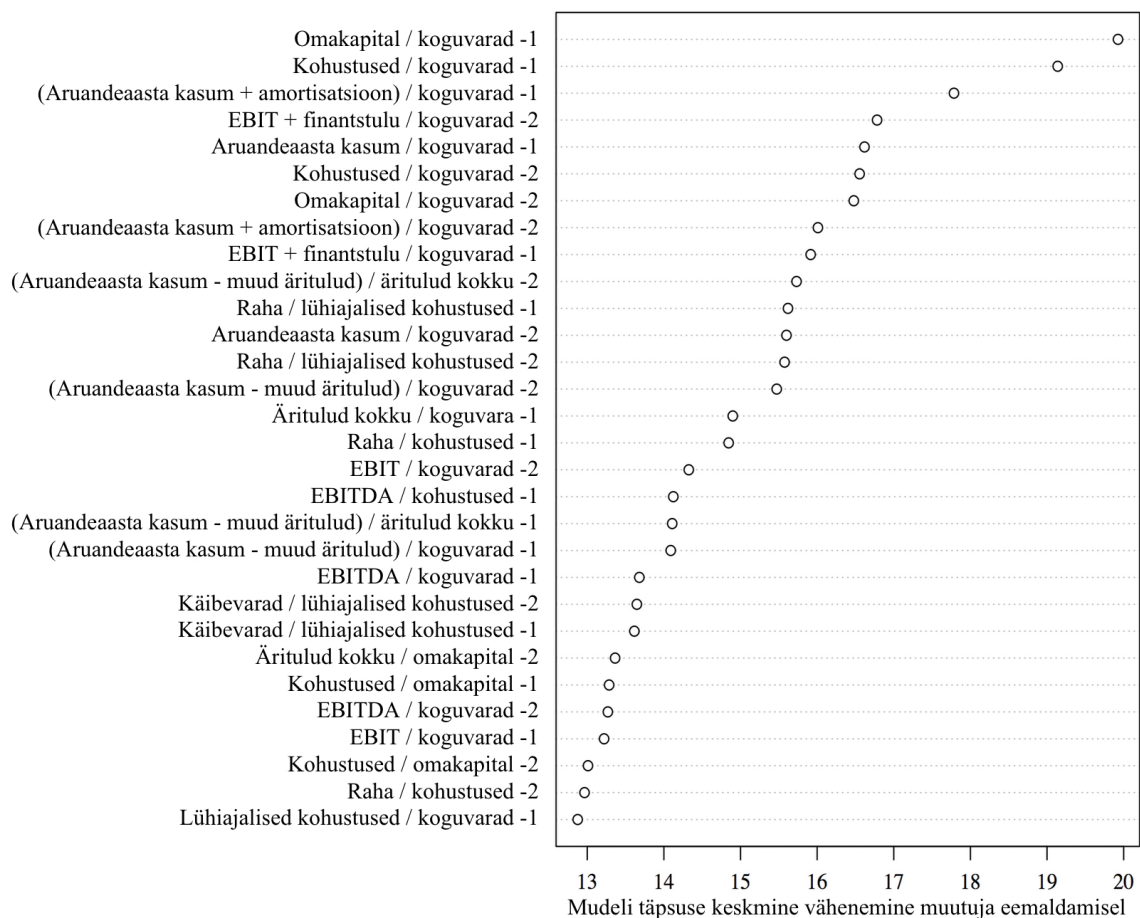
Joonis 2.3 Juhusliku metsa klassifikatsioonivea muutus (kotist välja meetodil ehk *out of the bag error*) genereeritud klassifikatsioonipuude arvu kasvades.

Lihtsa joonise abil on võimalik hinnata, millisest hetkest ehk siis genereeritud puude arvust alates tekib juhusliku metsa meetodit kasutades tulemuste stabiliseerumine ning

milline võiks olla antud meetodi puhul võimalik saavutatav klassifikatsioonimäär. Klassifikatsioonimäära arvutamine *kotist välja* (*out of the bag*) meetodil käib iga puu puhul nende vaatluste põhjal, mida antud puu kingapaela (*bootstrapping*) meetodil treenimiseks ei kasutatud. Ühe puu treenimiseks kasutatakse $\sim 2/3$ valimist ning kotist välja meetodi puhul kasutatakse kindla puu juures mudeli treenimiseks mittekasutatud osa valimist mudeli vea hindamiseks. Ehk sisuliselt on võimalik treeningvalimi andmetega saada suhteliselt lähedane tulemus sellele, mis saadakse mudeli kontrollil juhuslikul kontrollvalimil, mis pärineb üldkogumist. (Liaw, Wiener 2002: 18; Breiman 2001: 11) Antud töös on kontrollvalimi klassifikatsioonimäär ligilähedane sellele, mis joonisel 2.3 toodud. Seega saab väita, et kontrollvalim on juhuslikult valitud.

K-lähima naabri meetodi puuduseks, võrreldes nii logit mudeli kui ka juhusliku metsa meetodiga, oli selle puudus anda uurijale lisainformatsiooni muutujate olulisuse või mõju kohta hinnangute puhul, mis saadi mudeli rakendamisel. Juhusliku metsa puhul on aga mitmeid võimalusi muutujate olulisuse hindamiseks. Autor toob siin välja neist ühe ehk: keskmise täpsuse vähenemine (*Mean decrease accuracy*) joonisel 2.4 toodud muutuja eemaldamisel mudelist. Antud väärtused on arvutatud treeningvalimi korral ning andmestiku põhjal, mida konkreetsete klassifikatsioonipuude genereerimiseks ei kasutatud ehk siis jällegi “kotist välja” meetodil. Kõige olulisemaks ning sealt olulisuselt järgnevaks on muutujad, mille eemaldamisel mudelist üldine klassifikatsioonimäär kõige rohkem väheneb (Breiman 2001: 23-25). Joonise 2.4 puhul vajab märkimist, et see kehtib ainult ja ainult selle muutujate komplekti kasutamise korral. Klassifikatsioonimäära vähenemine on toodud vaid ühe muutuja eemaldamisel ning teiste samaks jäädes.

Mudeli täpsust mõjutavateks kõige olulisemateks muutujateks on $t-1$ perioodi “omakapital / koguvarad”, $t-1$ perioodi “kohustused / koguvarad” ja $t-1$ perioodi “Aruandeaasta kasum + amortisatsioon / koguvarad”.



Joonis 2.4 Juhusliku metsa hinnangu keskmine täpsuse vähenemine muutujate eemaldamisel, muutujate tähtsuse järjekorras (autori koostatud).

Üllatavalt on keskmise täpsuse vähenemise pingereas kohtadel 4...7 kolm muutujat t-2 perioodist ehk kaks aastat varasemast ajast enne pankrotistumist ning üks samast perioodist (vt. tabel 2.4). Autori märkusena on pankrotistumise prognoosimise mudelites enamasti olulised kõige viimase perioodi muutajad ning see on siinkohal huvipakkuv olukord. Nendeks muutujateks on: t-2 perioodi “EBIT + finantstulu / koguvarad”, t-1 aasta “Aruandeaasta kasum / koguvarad”, t-2 perioodi “Kohustused / koguvarad” ja t-2 perioodi “omakapital / koguvarad”. Kui enamasti on olulised t-1 perioodi muutujad, siis autori meelest on tähelepanuväärne, et juhusliku metsa puhul on ka t-2 perioodi muutujad omandanud nii suure olulisuse pingereas. Pigem eeldas autor, et t-2 perioodi muutujad oleksid tähtsuselt pingereas selgelt tagapool t-1 perioodi näitajatest.

Tähtsusest esimese 7 olulisema muutuja puhul oli toodud nii t-1 kui ka t-2 perioodist pärit “kohustused / koguvarad” ja “omakapital / koguvarad”, mis autori arvates näitab kui suurt mõju mõlemad suhtarvud autokaubanduse ettevõtete pankroti prognoosimise kohalt omavad. Eelnevatest aga tähtsuse järjekorras oli esmaselt kajastatud t-2 perioodi “EBIT + finantstulu / koguvarad” enne, kui vastav t-1 perioodi näitaja. See annab autorile alust arvata, et autokaubanduse ettevõtete puhul võib olla pankrotistumise protsess nende “EBIT + finantstulu / koguvarad” suhtarvu põhjal ennustatav pikemal kui üheaastasel perioodil. Muutused ebaõnnestunud ettevõtete intressi- ja maksueelse kasumi ning finantstulude summa põhiselt toimuvad varem kui üks aasta enne pankrotistumist. Antud eelduste kontrollimiseks tuleks täpsemalt modelleerida autokaubandusettevõtete pankrotistumise protsessi, mis võimaldaks saada autori eeldusele kinnituse või see ümber lükata.

Tabel 2.10 Juhusliku metsa klassifitseerimistäpsused sektorite lõikes, kontrollvalimi põhjal

Sektor:	G45.1		G45.2		G45.3		G45.4		G45	
Tegelik\mudel	0	1	0	1	0	1	0	1	0	1
0	327	22	263	21	130	17	37	8	757	68
1	39	281	28	157	23	123	2	41	92	602
Spetsiifilisus	93,70%		92,61%		88,44%		82,22%		91,76%	
Tundlikkus	87,81%		84,86%		84,25%		95,35%		86,74%	
Üldklassifitseerimine	90,88%		89,55%		86,35%		88,64%		89,47%	

Allikas: Autori koostatud.

Mudeli klassifitseerimismäärade stabiilsus (tabel 2.10) kõigi alamsektorite lõikes v.a. alamsektor G45.3, annab koostajale kindluse, et see töötab ka juhuslikul antud sektorist pärit kontrollvalimil. Sektori G45.3 madal tundlikkus tuleb ilmselt osavaliimi väiksusest ning ilmselt on ka osalt lihtsalt juhuslikkuse komponendist sõltuv.

Juhusliku metsa klassifitseerimist kasutades oli autori jaoks lõplik klassifitseerimistäpsus, spetsiifilisus ning tundlikkus suhteliselt ootuspärased võrrelduna näiteks tulemustega, mis saadud kasutades erinevaid klassifikatsioonipuid. Arvestades käesoleva sektori tegevusalade suhtelist heterogeensust, siis paremat klassifitseerimistäpsust kontrollvalimil on raske loota. Samas oli üllatav see, et mudeli puhul

omandasid nii suure olulisuse kolm t-2 perioodi muutujat kuuludes tähtsusest esimese 15% muutujate hulka.

2.4. Erinevatel modelleerimise meetoditel saadud tulemuste võrdlus

Autori eeldus kolme erinevat mudelit koostades sai selle töö käigus kinnitust. Klassikalised statistilised mudelid ning nende klassifikatsioonimäär jääb alla masinõppimise meetoditele. Siinkohal küll on see ainult kahe näitega toodud, kuid leidis selle töö raames kinnitust. Andmete jaotamine kaheks osaks ehk treening- ning kontrollvalimiks oli masinõppimise meetodi paremaks kontrolliks vajalik nõue, kuna vastasel juhul võivad masinõppimise meetodid anda treeningvalimil 100% klassifitseerimistäpsuse, ent seda peab kontrollima, et oleks võimalik mudeli kohta järeldusi teha.

Masinõppimise meetodite puhul pole olulised klassikaliste meetodite puhul muutujatele esitatud nõuded: normaaljaotuse eeldus, multikollineaarsuse puudumine muutujate komplektist, multimodaalsuse puudumine ja muutujate suurem koguarv vaatlustest. Seega võimaldavad need meetodid andmekaota analüüsida ettetulevaid situatsioone. Antud töös kasutatud juhusliku metsa meetodi puhul pole vaja andmetest eemaldada või tasandada ekstreemumeid, kuna see pole arvutusvalemis oluline. Teisalt k-lähima naabri meetodi puhul võivad ekstreemväärtused andmestikus kahandada klassifikatsiooni määra, kuna klassifitseerimise aluseks on vaatluste geomeetriline kaugus üksteisest n -mõõtmelises ruumis, milles n on defineeritud muutujate arvuga.

Tulles tagasi mudeli klassifikatsioonimäärade juurde, siis logit mudeli puhul oli kontrollvalimi klassifikatsioonimäär 77,75%, mida võib võrdluses antud uurimisvaldkonna teiste logit mudelitega pidada suhteliselt kesiseks tulemuseks. Autor muutis mudeli tõenäosusskooride 0...1 poolitamiseks kasutatavat lõikepunkti 0,5 ning võttis kasutusele lõikepunkti 0,435, mille korral mudeli tundlikkus tõusis ning kontrollvalimi puhul on spetsiifilisus peaaegu võrdne tundlikkusega. Muu lõikepunkti kasutamine tavapärase 0,5 asemel on autori hinnangul vajalik, kuna mudelisse hindamiseks

kaasatud grupid pankrotistunud ja aktiivsetest ettevõtetest polnud sama suured, vaid pankrotistunud ettevõtteid oli 694 ning aktiivseid 825. Seetõttu on mudeli 0...1 hinnangute väärtuste sagedus suurem aktiivsete ettevõtete poolel ehk väärtusele 0 lähemal ning seetõttu omab 0,5st väiksemaid väärtusi hulk pankrotistunud ettevõtteid. Seetõttu muutis autor siin presenteerimiseks lõikepunkti väärtust, kuni saavutati kõige parem üldine klassifikatsiooni-määr treeningvalimi põhjal. (vt Tabel 2.11)

Tabel 2.11 Klassifikatsioonitabel logit mudel, k-lähima naabri meetod ning juhuslik mets, treening- ning kontrollvalim

Treeningvalim:	Logit mudel lõikepunkt 0.435	K-lähima naabri meetod (k=3)	Juhuslik mets (n=700)
Õiged pankrot	544	550	673
I liiki viga	129	31	0
Õige aktiivne	675	815	846
II liiki viga	171	123	0
Kokku:	1519	1519	1519
Spetsiifilisus	79,79%	96,34%	100,00%
Tundlikkus	80,83%	81,72%	100,00%
Klassifikatsioonimäär	80,25%	89,86%	100,00%
Kontrollvalim:	Logit mudel lõikepunkt 0.435	K-lähima naabri meetod (k=3)	Juhuslik mets (n=700)
Õiged pankrot	538	517	602
I liiki viga	156	50	68
Õige aktiivne	643	775	757
II liiki viga	182	177	92
Kokku:	1519	1519	1519
Spetsiifilisus	77,94%	93,94%	91,76%
Tundlikkus	77,52%	74,50%	86,74%
Klassifikatsioonimäär	77,75%	85,06%	89,47%

Allikas: Autori koostatud.

Masinõppimise meetodid andsid paremaid tulemusi kui logit mudel. Näiteks k-lähima naabri meetodi korral on kontrollvalimi klassifikatsioonimääraks 85.06%, mille puhul valmistab autorile muret mudeli kesine tundlikkus ehk mudel ei suuda väga hästi klassifitseerida pankrotistuvaid ettevõtteid, mis peaks olema pankrotistumise prognoosimise mudelite koostamisel peamiseks eesmärgiks. (vt Tabel 2.11) Juhusliku metsa meetodit kasutades on seega kontrollvalimi klassifitseerimistäpsus suurim. Ka mudeli tundlikkus ja spetsiifilisus on suhteliselt võrdsed. Antud mudeli klassifikatsiooni

määr, arvestades valimi koostamise printsiipe ehk valim on koostatud juhuslikult ning välditud kunstlikku sobitamist, on äärmiselt hea. (vt Tabel 2.11)

K-lähima naabri meetodi puhul ei ole võimalik hinnata mudelisse kaasatud sõltumatute muutujate olulisust, kuid see võimalus on logit mudeli puhul, mille korral on autor (vt. tabel 2.12) toonud mudeli hindamiseks kasutatud 4 kõige olulisemat muutujat ning Juhusliku metsa meetodi puhul on toodud olulisuse põhjal 14 järjestatud muutujat kõige olulisematest, mille eemaldamine mudelist tooks kaasa kõige suurema klassifikatsiooni-täpsuse languse. Ainuke muutuja, mis kattus mõlema meetodi puhul, oli omakapitali suhe koguvaradesse. Kui juhusliku metsa meetodi puhul oli see kõige tähtsam muutuja (vt joonis 2.4), siis logit mudeli jaoks tähtsuselt teine muutuja (vt tabel 2.7).

Tabel 2.12 Logit mudeli muutujad ning Juhusliku metsa meetodi puhul kõige olulisemad 14 muutujat

Logit mudel:		
Muutuja nimi	Arvutusvalem	Periood
kapstr4_1	(Varud + Nõuded ja ettemaksed) / koguvarad	t-1
fin4_1	Omakapital / koguvarad	t-1
liq2_1	Raha / koguvarad	t-1
rent1_1	(Aruandeaasta kasum - muud äritulud) / (müügitulu + muud äritulud)	t-1
Juhusliku metsa meetod:		
Muutuja nimi	Arvutusvalem	Periood
fin4_1	Omakapital / koguvarad	t-1
fin2_1	Kohustused / koguvarad	t-1
ef2_1	(Aruandeaasta kasum + amortisatsioon) / koguvarad	t-1
rent7_2	(EBIT + finantstulud) / koguvarad	t-2
rent9_1	Aruandeaasta kasum / koguvarad	t-1
fin2_2	Kohustused / koguvarad	t-2
fin4_2	Omakapital / koguvarad	t-2
ef2_2	(Aruandeaasta kasum + amortisatsioon) / koguvarad	t-2
rent7_1	(EBIT + finantstulud) / koguvarad	t-1
rent1_2*	(Aruandeaasta kasum - muud äritulud) / (müügitulu + muud äritulud)	t-2
makse2_1	Raha / lühiajalised kohustused	t-1
rent9_2	Aruandeaasta kasum / koguvarad	t-2
makse2_2	Raha / lühiajalised kohustused	t-2
rent3_2*	Aruandeaasta kasum - muud äritulud / koguvarad	t-2

Allikas: autori arvutused.

Käesoleva töö koostamisel oli autoripoolseks suurimaks üllatuseks juhusliku metsa meetodi puhul esimese 14 olulisema muutuja hulka 59st mudeli treenimiseks kasutatud

muutujast ehk 24% kõige olulisemate muutujate hulka kuulusid 8 t-2 perioodist pärit näitajat. Nendeks olid: “rent7_2” ehk EBIT + finantstulu / koguvarad, “fin2_2” ehk kohustused / koguvarad, “fin4_2” ehk omakapital / koguvarad, “ef2_2” aruandeaasta kasum + amortisatsioon / koguvarad, “rent1_2” ehk (Aruandeaasta kasum - muud äritulud) / (müügitulu + muud äritulud), “rent9_2” ehk aruandeaasta kasum / koguvarad, “makse2_2” ehk raha / lühiajalised kohustused ja “rent3_2” ehk aruandeaasta kasum - muud äritulud / koguvarad. Olles põhjalikult tutvunud antud valdkonna põhiste uuringutega, siis võib tõdeda, et väga harva on olulist mõju mudelis omanud vanemad kui t-1 perioodi andmed.

Esimesest 14 olulisemast muutujast nende eemaldamisel mudelist üldise klassifikatsioonimäära langemisel juhusliku metsa meetodil moodustub 6 muutujate paari, millesse kuuluvad nii esimese kui teise perioodi vastav finantssuhtarv. Nendeks muutujateks on fin4, fin2, ef2, rent7, rent9 ja makse2. Autori arvates on need juhusliku metsa meetodi puhul kõige olulisemateks muutujateks. Siinkohal tuleb arvestada sellega, et antud muutujate toodud olulisuse järjestus kehtib ainult vastava mudeli treenimisel kaasatud muutujate komplekti juures. Selle järjestuse puhul on oluline ka teiste mudelisse kaasatud muutujate koosmõju.

Kokkuvõtvalt osutusid kõige olulisemaks autokaubanduse sektori pankroti prognoosimise seisukohalt 2 finantsstruktuuri kirjeldavat näitajat ehk omakapital / koguvarad, mis pankrotistunud ettevõtete keskmisena oli negatiivne ning kohustused / omakapital, mille puhul suurem aktiivsete ettevõtete puhul keskmine alla ühe ehk kohustused olid väiksemad omakapitalist ning pankrotistunud ettevõtete puhul oli keskmine suhtarvu tase üle ühe, mille puhul kohustuste maht ületas omakapitali mahtu. Eelnevalt kajastatud näitajate puhul olid olulised nii t-1 kui ka t-2 perioodist pärit olevad näitajad Oluline oli perioodist t-2 Aruandeaasta kasum + amortisatsioon / koguvarad, mis oli mudelis oma tähtsusest kolmas näitaja, mis annab autorile aluse arvata, et keskmiselt toimub autokaubandusettevõtete tulemuste halvenemine juba rohkem kui üks aasta ja 6 kuud enne pankrotistumist. Järgnesid rentaablust kajastavad näitajad rent7 ning rent9 mõlemast perioodist pärit suhtarvud, mis omasid suurt mõju.

Kuid lisaks erinevate mudelite klassifikatsioonitäpsustele on oluline kajastada võimalikke I ning II liiki vigade hinda. Esimese puhul ehk I liiki vea (vale positiivne) puhul tehakse otsus, et aktiivne ettevõtte pankrotistub, kuigi tegelikkuses mitte, ning selle vea hinnaks on saamata jäänud tulu. Teise puhul siis II liiki viga (vale negatiivne), mille puhul kannab krediitor teatud osa väärtuses laenusummast kahju, kuna see jääb tagasi maksmata. See annab veel ühe dimensiooni mudelite kvaliteedi hindamiseks, sest tehtud vigade hind ei pruugi üksüheselt korreleeruda klassifikatsioonitäpsusega ning sellest on tihtilugu pankrotiproгноosimise mudelite koostamisel mööda vaadatud.

Seega autor leiab, et antud töö edasiarenduseks oleks mõistlik kasutada juhusliku metsa meetodit, kuna see andis kõige täpsema klassifikatsiooni määra. Lisaks võimaldab see meetod arvutada tõenäosusskoorid kõigile valimis olevatele ettevõtetele. Järgneva sammuna oleks oluline teada saada kui suur võib olla keskmine krediitori poolt jäämata saanud tulu laenu mitteandmisel ning kui suur on keskmine kulu mis pangalt tuleb maha kanda ettevõtte pankrotistumisel. Eelneva alusel oleks võimalik arvutada rahalises mõõtmes murdepunktid mudelitele, mille korral oleks krediitori laenu mitte tagasimaksmisel saamata jäänud tulu ning laenu andmata jätmisel saamata jääv tulu minimaalne.

KOKKUVÕTE

Töö autor seadis eesmärgiks koostada käesolevas magistritöös Euroopa autokaubanduse sektori ettevõtete pankrotistumise prognoosimise mudelid. Töös rakendati klassikalise statistilise meetodina logit mudelit ja masinõppimise meetoditena k-lähima naabri ning juhusliku metsa meetodit.

Esmalt käsitles töö autor ettevõtete ebaõnnestumise ja pankrotistumise mõistet varasemate uurimuste näitel ning tõi välja kuidas on varasemalt kaubandussektorile keskendunud ebaõnnestumise või pankrotistumise prognoosimise töodes seda tehtud. Shirata (1998) mudel keskendus ebaõnnestumise prognoosimisele, siis vastukaaluks antud töös kajastatud töödest valdav enamus ehk Bhargava ja teised (1998), McGurr ja Devaney (1998), Lukason (2006), Lee ja Choi (2013) ning käesoleva magistritöö autor enda bakalaureuse töös (2013) kajastasid ettevõtte ebaõnnestumisena selle kõige tõsisemat vormi ehk pankrotistumist. Sellest sai antud töö autor kindluse koostada ettevõtete pankrotistumise prognoosimise mudel.

Varasemalt kasutatud pankrotistumise prognoosimise meetodite valik on suhteliselt lai ning võib jagada meetodid kahte suuremasse kategooriasse ehk klassikalised statistilised meetodid ning masinõppimise põhised meetodid. Esimeste hulka kuuluvad enimkasutatutena nii mitmene diskriminantanalüüs kui ka logit mudel, millest viimast rakendas autor enda bakalaureusetöös. Masinõppimise põhiste meetoditena kajastati neist kõige populaarsemaid ning ka kahte vähemkasutatud meetodeid, mis selles töös autori poolt rakendati ehk k-lähima naabri meetodit ning juhusliku metsa meetodit.

Seejärel koostas autor ülevaate kaubanduse sektori pankroti prognoosimisel senises kirjanduses kajastatud olulistest finantssuhtarvudest jagades need vastavatesse ettevõtte majandustegevuse aspekte kirjeldavatesse kategooriatesse: lühiajalise maksevõime,

efektiivsuse, rentaabluuse, finantsstruktuuri ja varade struktuuri suhtarvud ning finantseerimisallikate dünaamika näitajad. Varasemates töödes on enim kordi oluliseks osutunud rentaabluuse ning ettevõtte toimumise efektiivsuse suhtarvud.

Automüügi sektori ettevõtted moodustavad umbes 13% kõigist kaubandussektori ettevõtetest. Nende osakaal moodustab Euroopa ettevõtte koguarvust 3,6%. Kaubandussektori käive moodustas 2013. aastal Euroopa liidu kõikide ettevõtte käibest 37,5%. Autokaubanduse sektor moodustas samal aastal umbes 4% kogu ühenduse alal tegutsenud ettevõtte käibest (Eurostat 2016). Mootorsõiduk on tarbeese, kuid töös toodud dünaamika põhjal võib järeldada, et mootorsõiduk soetamine lükatakse edasi negatiivse majandusväljavaate puhul, mis tähendab, et antud sektori müügikäive on volatiilsem kui majandus tervikuna, mis tähendab et autokaubanduse sektori ettevõtted peavad olema valmis toime tulema müügitulemuste võnkumisega aastate lõikes.

Klassikalise statistilise meetodina rakendas autor antud töös logit mudelit. Mudel klassifitseeris murdepunkti 0,435 juures kontrollvalimis olevatest ettevõtetest õigesti 77,75%. Antud tulemus on logit mudeli puhul pigem keskmine, kuid arvestades andmestikku kuuluvate alamsektorite ettevõtte struktuuride suhteliselt suurt heterogeensust, siis on autori arvates antud tulemus rahuldav. Logit mudel on nimelt suure üldistusvõimega, kuid seetõttu pole võimalik luua keerukamate andmestruktuuride jaoks sobivaid mudeleid.

Logit mudeli üldkuju on toodud järgnevalt (kus K_1 on defineeritud pankrotistumise

tõenäosuse avaldises $p = \frac{I}{I + e^{-K_1}}$):

$$K_1 = -5,272 \frac{\text{raha}}{\text{koguvarad}} - 5,525 \frac{\text{aruandeaasta kasum} - \text{muud äritulud}}{\text{äritulud kokku}} - 2,873 \frac{\text{omakapital}}{\text{koguvarad}} - 1,029 \frac{\text{varud} + \text{nõuded ja ettemaksed}}{\text{koguvarad}} + 0,514$$

Logit mudeli puhul mõjutas pankrotistumise tõenäosust enim muutuja (varud + nõuded ja ettemaksed) / koguvarad üheühikuline kasv, mille tulemusena väheneb pankrotistumise tõenäosus 37,5% võrra.

Teise meetodina rakendas autor antud töös k-lähima naabri meetodit (*k-nearest-neighbor discriminant analysis*), mis liigitub masinõppimise meetodite alla. Antud meetodi puhul toimub klassifitseerimine n-mõõtmelises ruumis Eukleidese distantssi alusel, milles n on defineeritud muutujate arvuga. Seetõttu on see tundlik erineva suurusjärguga muutumispirkonnaga muutujate suhtes. Parima tulemuse antud meetodi puhul andis parameetri k väärtuseks valitud 3, mille puhul oli õigesti klassifitseerimise määr kontrollvalimil 85,06%. Antud meetodit rakendades oli mudeli tundlikkus kontrollvalimil suhteliselt madal jäädes 74,5% juurde. See tähendab, et kontrollvalimi puhul suutis mudel pankrotistunud ettevõtetest õigesti klassifitseerida vaid alla 3/4.

Kolmanda meetodina rakendas käesolevas magistritöös juhusliku metsa metsa meetod (*random forest*), mis oma olemuselt on klassifikatsiooni ja regressiooni puude kogum, mille põhjal toimub vaatluste klassifitseerimine. Meetodi puhul on üheks parameetriks puude arv, mis mudeli koostamisel genereeritakse. Puude arvu kasvades toimub klassifikatsioonitäpsuse paranemine ning stabiliseerumine seni, mil mudeli koostaja hinnangul mudel saavutab stabiilsuse ning muutub ebavajalikuks metsa kasvatamine, kuna see ei panusta klassifikatsioonimäära tõusu. Antud juhul toimus klassifikatsioonitäpsuse stabiliseerumine 700 genereeritud klassifikatsioonipuu juures. Juhusliku metsa meetodi puhul kasutatakse puude genereerimisel nn “kingapaela” meetodit, mille puhul $\sim 2/3$ juhuslikult valitud osa valimist kasutatakse puu genereerimiseks ning ülejäänut kontrollvalimina. Juhuslikult valitud $2/3$ muutub iga järgneva puu genereerimisel.

Autor kasutas juhusliku meetodil mudeli treenimiseks 29 “t-1” perioodist pärit muutujat ning sama nomenklatuuriga (29) “t-2” perioodist pärit finantssuhtarvu ning lisaks ühte suurust kajastavat näitajat ehk koguvarade mahtu. See teeb kaasatud sõltumatute muutujate koguarvuks 59. Mudeli klassifikatsioonitäpsus treeningvalimi puhul oli 100%, mis on treeningvalimi mahu ning sellesse kuuluva info hulga piisavusel tavapärane tulemus ning seetõttu ei ole mõtet treeningvalimi kõrget klassifikatsioonimäära selle meetodi puhul üle tähtsustada. Kontrollvalimi puhul on klassifikatsioonimäär 89,47%, mis on arvestades antud sektorisse kuuluvate ettevõtete heterogeensust autori hinnangul väga hea tulemus. Mudeli tundlikkus oli 86,74% ning

spetsiifilisus 91,76%. Vastav tundlikkuse näitaja oleks autori hinnangul võinud olla spetsiifilisuse arvelt kõrgem olla.

Magistritöö tulemusena võib autor väita, et juhusliku metsa meetod pankrotistumise prognoosimiseks annab häid tulemusi ning seega soovitab antud magistritöö autor seda pankrotistumise prognoosimisel tulevikus rakendada. Meetodi puhul on selle rakendamisel eeliseks selle lihtsus ning ka saadava lisainformatsiooni maht.

Autokaubanduse sektori pankroti prognoosimisel osutusid juhusliku metsa meetodit rakendades kõige olulisemateks muutujateks mudelist eemaldamisel tekkiva klassifikatsioonitäpsuse vähenemise mõistes järgnevad kaks finantsstruktuuri kirjeldavat suhtarvu ehk omakapital / koguvarad ja kohustused / omakapital. Eelnevalt kajastatud näitajate puhul olid olulised nii t-1 kui ka t-2 perioodist pärit näitajad. Samade kriteeriumite põhjal osutus oluliseks perioodist t-2 (aruandeaasta kasum + amortisatsioon) / koguvarad, mis oli mudelis oma tähtsusest kolmas näitaja, mis annab autorile aluse arvata, et autokaubandusettevõtete tulemuste halvenemine toimub juba rohkem kui üks aasta ja 6 kuud enne pankrotistumist. Järgnesid rentaablust kajastavad näitajad (EBIT + finantstulu) / koguvarad ja aruandeaasta kasum / koguvarad, mille mõlema puhul olid kajastatud t-1 kui ka t-2 periood, mis omasid suurt mõju mudeli klassifikatsioonitäpsusele.

Töö edasiarenduseks eelnevalt kirjeldatu põhjal on mõistlik kasutada juhusliku metsa meetodit, kuna seda rakendades saadi selle töö puhul kõige kõrgem õigesti-klassifitseerimise määr. Pankrotistumise prognoosimise vallas krediidiasutuste seisukohalt ei oma kõige suuremat olulisust mudelite juures mitte õigesti-klassifitseerimise määr ega ka kõrge mudeli tundlikkus vaid pigem saamata jäänud tulu ja laenude tagasimaksmata jätmisest kannatatava kahju minimeerimine. Autor leiab, et antud töö arendusena võiks koostada pankrotistumise prognoosimise mudeli, mille optimeerimine toimuks kahjude minimeerimise aspektist lähtuvalt. Selleks on vaja koostada lisaks praegusele kahjude mahtu hindav mudel, millega saaks liita selles töös koostatud pankrotistumise prognoosimise mudeli.

VIIDATUD ALLIKAD

1. **Adnan Aziz, M., Dar, A.** Predicting corporate bankruptcy: where we stand? *Corporate Governance*, 2006, Vol. 6, No. 1, pp. 18-33.
2. **Altman, E. I.** Financial Ratios, Discriminant Analysis and The Prediction of Corporate Bankruptcy. *The Journal of Finance*, 1968, Vol. 23, No. 4, pp. 589–609.
3. **Altman, E.I., Narayanan, P.** An international survey of business failure classification models. *Financial Markets, Institutions & Instruments*, 1997, Vol 6, No. 2, pp. 1-57.
4. **Altman, E.I.** Corporate Financial Distress and Bankruptcy: A Complete Guide to Predicting and Avoiding Distress and Profitting from Bankruptcy. Aufl., New York , 1993.
5. **Andreica, M.E., Cristescu, A., Stanila, L., Cataniciu, N.** Comparative analysis of early warning models for distressed firms.
6. **Balcaen, S., Ooghe, H.** 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems. *The British Accounting Review*, 2006, Vol. 38, pp. 63-93.
7. **Balcaen, S., Ooghe, H.** Alternative methodologies in studies on business failure: do they produce better results than the classical statistical methods. *Vlerick Leuven Gent Management School Working Papers Series*, 2004, pp. 1-33.
8. **Battacharya, S., Gepp, A., Kumar, K.** Business failure Prediction using Decision Trees. *Journal of Forecasting*, 2010, Vol. 29, pp. 536–555.
9. **Beaver, W.** Financial ratios predictors of failure. Empirical research in accounting: selected studies 1966. *Journal of Accounting Research*, 1967, Vol. 4, pp. 71-111.
10. **Bhargava, M., Dubelaar, C., Scott, T.** Predicting bankruptcy in the retail sector: an examination of the validity of key measures of performance. *Journal of Retailing and Consumer Services*, Vancouver, 1998, Vol. 5, No. 2, pp.105-117.

11. **Bilderbeek, J.** An empirical study of the predictive ability of financial ratios in the Netherlands. *Zeitschrift für Betriebswirtschaft*, 1979, Vol. 5, pp. 388-407. Viidatud Bilderbeek, J., Pompe. P. P. M. The prediction of bankruptcy of small- and medium-sized industrial firms, *Journal of Business Venturing*, 2005, Vol. 20, pp. 847–868 vahendusel.
12. **Bilderbeek, J., Pompe. P. P. M.** The prediction of bankruptcy of small- and medium-sized industrial firms. *Journal of Business Venturing*, 2005, Vol. 20, pp. 847-868.
13. **Breiman, L.** 2001. Random forests. *Machine learning*, 2001, Vol. 45, No. 1, pp. 5-32.
14. Bureau of Business Research. A Test Analysis of Unsuccessful Industrial Companies, *Bulletin Urbana: University of Illinois Press*, 1930, No. 31.
15. **Casey, C. and Bartczak, N.** 1985. Using operating cash flow data to predict financial distress: Some extensions. *Journal of Accounting Research*, pp.384-401.
16. **Daubie, M., Levecq, P., Meskens, N.** A comparison of the rough sets and recursive partitioning induction approach: An application to commercial loans. *International Transactions in Operational Research*, 2002, Vol. 9, pp. 681-694.
17. **Deakin, E.B.** A discriminant analysis of predictors of business failure. *Journal of accounting Research*, 1972, pp.167-179.
18. **Dimitras, A.I., Zanakis, S.H., Zopounidis, C. A.** Survey of Business Failures with an Emphasis on Prediction Methods and Industrial Applications. *European Journal of Operational Research*, 1995, No. 90, pp. 487-513.
19. **du Jardin, P.** The Influence of Variable Selection Methods on the Accuracy of Bankruptcy Prediction Models. *Bankers, Markets & Investors*, 2012, Vol. 116.
20. Eurostat. European Commission. [<http://ec.europa.eu/Eurostat>] 12.05.2016.
21. **Ghio, E.** Coping with Business Failure and Bankruptcy in the European Union. A Case Study on the Failure of Market Integration. A Case Study on the Failure of Market Integration 2015.

22. **Gilbert, L.R., Menon, K., Schwartz, K.B.** Predicting bankruptcy for firms in financial distress. *Journal of Business Finance & Accounting*, 1990, Vol. 17, No. 1, pp.161-171.
23. **Han, I., Jo, H., Lee, H.** Bankruptcy Prediction Using Case-Based Reasoning, Neural Networks, and Discriminant Analysis. *Expert Systems With Applications*, 1997, Vol. 12, No. 2., pp. 97-108.
24. **Harrell, F. E.** Regression modelling strategies: with applications to linearmodels, logistic regression, and survival analysis. New York: Springer, 2001, pp. 352.
25. **He, Y. and Kamath, R.** Business failure prediction in retail industry: an empirical evaluation of generic bankruptcy prediction models. *Academy of Accounting and Financial Studies Journal*, 2006, Vol. 10, No. 2, pp. 97.
26. **Horrigan, J. O.** A History of Financial Ratio Analysis. *Accounting Review*, 1968, Vol. 43, No. 2, pp. 284-294.
27. **Hsieh, S.** A note on the optimal cutoff point in bankruptcy prediction models, *Journal of Business Finance and Accounting*, 1993, Vol. 20, pp. 457-464.
28. **John, G. H., Kohavi, R., Pfleger, K.** Irrelevant Features and the Subset Selection Problem, in *Machine Learning: Proceedings of the 11th International Conference*, New Brunswick, New Jersey, 1994, Vol. 15, No.10, pp. 121-129.
29. **Kahya, E.** Prediction of business failure: A funds flow approach. *Managerial Finance*, 1997, Vol. 23, No. 3, pp. 64-71.
30. **Kiang, M.Y.** A comparative assessment of classification methods. *Decision Support Systems*, 2003, Vol. 35, No. 4, pp. 441-454.
31. **Kumar, P.R., Ravi, V.** Bankruptcy prediction in banks and firms via statistical and intelligent techniques—A review. *European journal of operational research*, 2007, Vol. 180, No. 1, pp. 1-28.
32. **Kwak, W., Cheng, X., Ni, J.** Predicting bankruptcy after the Sarbanes-Oxley Act using logit analysis. *Journal of Business & Economics Research (Online)*, 2012, Vol. 10, No. 9, p. 521.
33. **Kwon, W.J.** The role of insurance in corporate risk finance. *Review of Business*, 2003, Vol. 24, No. 3, p. 36.

34. **Laitinen, E. K.** Financial Ratios and Different Failure Processes. *Journal of Business Finance & Accounting*, 1991, Vol. 18, No. 5, pp. 649-673.
35. **Laitinen, E. K.** The Duality of Bankruptcy Process in Finland. *The European Accounting Review*, 1995, Vol. 4, No. 3, pp. 433-454.
36. **Laitinen, E.K., Laitinen, T.** Misclassification in bankruptcy prediction in Finland: Human information processing approach. *Accounting, Auditing & Accountability Journal*, 1998, Vol. 11, No. 2, pp. 216-244.
37. **Laitinen, T., Sere, K., van Wezel, M.** Choosing bankruptcy predictors using discriminant analysis, logit analysis and genetic algorithms. 1996.
38. **Lee, S., Choi, W. S.** A multi-industry bankruptcy prediction model using back-propagation neural network and multivariate discriminant analysis. *Expert Systems with Applications*. 2013, Vol. 40, No. 8, pp. 2941-2946.
39. **Liaw, A. and Wiener, M.** Classification and regression by randomForest. *R news*, 2002, Vol. 2, No. 3, pp. 18-22.
40. **Lukason, O.** Pankrotistumiste modelleerimine Eesti kaubandusettevõtete näitel. TÜ Rahanduse ja arvestuse instituut, 2006, 79 lk. (magistritöö)
41. **Mcgurr, P.T., Devaney, S.A.** A retail failure prediction model. *The International Review of Retail, Distribution and Consumer Research*, 1998, Vol. 8, No. 3, pp. 259-276.
42. **Muru, L.** Kindlustuskahjude sageduse analüüs lokaalse regressiooni ja k-lähima naabri meetodil. TÜ Matemaatilise statistika instituut, 2015, 53 lk. (magistritöö)
43. **O'Brien, R. M.** A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Quality & Quantity*, 2007, Vol. 41, pp. 673-690.
44. **Ohlson, J. A.** Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, 1980, Vol. 18, No. 1, pp. 109-131.
45. **Ooghe, H., Spaenjers, C. Vandermoere, P.** Business Failure Prediction: Simple Intuitive Models Versus Statistical Models. *The IUP Journal of Business Strategy*, 2009, Vol. 6, No. 3/4, pp. 7-44.
46. **Paal, M.** Pankroti prognoosimise mudeli koostamine Eesti kaubandussektori ettevõtete näitel. TÜ ettevõtte majanduse instituut, 2013, 57lk. (bakalaureusetöö)

47. **Pervan, I., Pervan, M., Vukoja, B.** Prediction of company bankruptcy using statistical techniques—Case of Croatia. *Croatian Operational Research Review*, 2011. Vol. 2, No. 1, pp. 158-167.
48. **Rafiei, F. M., Manzari, S. M., Bostanian, S.** Financial health prediction models using artificial neural networks, genetic algorithm and multivariate discriminant analysis: Iranian evidence. *Expert Systems with Applications*. 2011, Vol. 38, No. 8, pp. 10210-10217.
49. **Ropega, J.** The Reasons and Symptoms of Failure in SME. *Int Adv Econ Res*, 2011, Vol. 17, pp. 476-483.
50. **Scott, J.** The probability of bankruptcy: a comparison of empirical predictions and theoretical models. *Journal of Banking and Finance*, 1981, Vol. 5, No. 3, pp. 317-344.
51. **Shirata, C.Y.** Financial ratios as predictors of bankruptcy in Japan: an empirical research. *Tsukuba College of Technology Japan*, 1998, pp. 1-17.
52. **Slavici, T., Maris S., Pirtea, M.** Usage of artificial neural networks for optimal bankruptcy forecasting. Case study: Eastern European small manufacturing enterprises. *Quality & Quantity*. 2016, Vol. 50, No. 1, pp. 385-398.
53. **Stahlman, R.** 2015. Üleeuroopalise tööstussektori mikro-ja väikeettevõtete pankroti prognoosimudeli klassifitseerimistäpsus erinevates ebaõnnestumise protsessides. TÜ ettevõtetemajanduse instituut, 2015, 136lk. (magistritöö)
54. **Theodossiou, P.** Alternative models for assessing the financial condition of business in Greece. *Journal of Business Finance & Accounting*, 1991, Vol. 18, No. 5, pp. 697-720.
55. **Theodossiou, T.** Predicting Shifts in the Mean of a Multivariate Time Series Process: An Application in Predicting Business Failures. *Journal of the American Statistical Association*, 1993, Vol. 88, pp. 441-449.
56. **Vuran, B.** Prediction of business failure: a comparison of discriminant and logistic regression analyses. *Istanbul University Journal of the School of Business Administration*, 2009, Vol. 38, No. 1, pp. 47-65.

57. **Wu, C., Ho, S.** Financial ratio adjustment: industry-wide effects or strategic management, *Review of Quantitative Finance and Accounting*, 1997, Vol. 9, pp. 71-88.
58. **Wu, W.W.** Beyond business failure prediction. *Expert systems with applications*, 2010, Vol. 37, No. 3, pp. 2371-2376.
59. **Zavgren, C.** The Prediction of Corporate Failure: The State of the Art. *Journal of Accounting Literature*, 1983, Vol. 2, pp. 1-37.
60. **Zheng, Q., Yanhui, J.** Financial Distress Prediction Based on Decision Tree Models. In *Service Operations and Logistics, and Informatics*, 2007. International Conference 2007, pp. 1-6.
61. Statistical Classification of Economic Activities in the European Community, Rev. 2 (2008) [http://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_NOM_DTL&StrNom=NACE_REV2&StrLanguageCode=EN&IntPcKey=18508904&StrLayoutCode=HIERARCHIC]. 08.04.2016

Summary

BANKRUPTCY PREDICTION MODELS OF PAN-EUROPEAN MOTOR VEHICLES WHOLESALE AND RETAIL TRADE BUSINESSES

Märt Paal

The field of business bankruptcy prediction (BBP), as we know it today, started with the Altman's (1968) article "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy" which introduced the Z-score. The majority of BBP models are focused on manufacturing sector but there are also a number of BBP models for mixed sectors. The fact that only six retail trade sector BBP models have been created prior to the current thesis (Bhargava et al. 1998, McGurr & Devaney 1998, Shirata 1998, Lukason 2006, Lee & Choi 2013, Paal 2013), indicating for a need to comprehensively explore the field for new improved models.

This master's thesis concentrates on creating bankruptcy prediction models for pan-european motor vehicles trade sector businesses. This sector has never been thoroughly studied in the field of BBP, meaning that this master's thesis could provide novel insight to the field. Additionally, new specific financial ratios could be used as unique bankruptcy predictors for the sector.

The aim of this thesis is to propose bankruptcy prediction models for pan-european motor vehicles trade sector using three different methods and concluding the study by finding which method has the highest classification accuracy. An associated task with the aim will be to find unique financial ratios that best discriminate bankrupt businesses from still successful. In order to reach the aim following tasks were created:

1. to clarify definitions of business failure and bankruptcy;

2. to give an overview of methods used in bankruptcy prediction;
3. to give an overview of financial ratios used in bankruptcy prediction studies, focusing on the trade sector;
4. to give a brief review on the European motor vehicles trade sector;
5. to build pan-european motor vehicles trade sector bankruptcy models using 3 different methods;
6. to analyse and compare the results of different methods.

In the first chapter the author gives an overview of the usage of business failure definitions in previous trade sector BBP studies. The definition of bankruptcy is mostly used for the studies focusing on trade sector. Classical statistic (numeric) and machine learning methods have been used in the BBP field. An overview of the most popular is given in this paper. The author also surveys the k-nearest neighbour method and random forest classification method, which are implemented for this thesis, even though relatively new methods in the field of BBP.

In the second chapter the author gives a brief overview of pan-european motor vehicle trade sector, including the revenue and the number of businesses dynamics for the period of 2010 to 2014. Motor vehicle trade sector's revenue constituted 4% of all of the business revenues combined in Europe in 2013 (Eurostat 2016).

The author of this thesis uses one classic statistical method and two machine learning methods for creating BBP models for this study. The first method used is the logit model, previously used by the author also to study (blablabla, mis asja sa siis uurisid) (Paal 2013). The second method is the K-nearest neighbour analysis for classification, and the third is the Random forest method. The latter two are slightly less frequently used in the BBP field but promising outcome, mainly concerning high classification accuracy for the control sample, is expected.

Numerical data used to create the statistical models is mined from the Bureau van Dijk Amadeus database. The dataset used in this paper consists of “wholesale and retail trade and repair of motor vehicles and motorcycles sector” (G45) balance sheet and income statement data. Author divides the selected data into two groups: learning and control sample to make conclusions on the utility of the machine learning methods. When using the machine learning methods the classification rate for training samples is often around 100% and usually greatly degrades for control sample. 1367 bankrupt businesses are included in the final analysis plus 1671 still active businesses.

The best logit model created for the motor vehicle trade sector had a classification rate of 77.75% at cutoff point 0.435 for the control sample. Cutoff point different from the default 0.5 was selected to maximise the overall classification rate. Giving that the logit model has a great ability for generalisation and is not prone to overfitting the data, this result is valid for the author. The most important variable for model described below is the “working capital / total assets”. The best logit model is following as (K_i is the score

in the following equation $p = \frac{1}{1 + e^{-K_i}}$):

$$K_i = -5.272 \frac{\text{cash}}{\text{total assets}} - 5.525 \frac{\text{profit}}{\text{total revenue}} - 2.873 \frac{\text{equity}}{\text{total assets}} - 1.029 \frac{\text{working capital}}{\text{total assets}} + 0.514$$

K-nearest neighbour discriminant is the second method applied in this thesis. This is regarded as a “black box” method because the only output from the analysis is classification of observations with no other statistics. An optimal value for parameter k was chosen to maximise the classification rate. The optimal k value 3 gave the 85.06% classification rate for the control sample. Sensitivity in this case is for the control sample is 74.5%, which is lower than the classification rate.

The third approach used in this paper is the random forest method. This method is based on a chosen number of regression trees. Number of trees is a free parameter chosen by the user, usually between a few hundred to a few thousand. In this case the converging of the classification rate was reached around 700 classification trees, which gave 100%

classification rate for the learning sample and OOB error being around 0.11. The classification rate for the control sample of this model was 89.47%, which in author's opinion is remarkable despite the difficult dataset consisting of businesses with heterogeneous financial structures. The sensitivity on this model was 86.74% and specificity was 91.76%. Author used 29 financial ratios from periods t-1 and t-2 and variable total from period t-1, meaning that the total number of variables used in generating the final and the best set of random forest is 59. The author is pleased with the classification results from the model created with the random forest method.

In conclusion, the author proposes that using the random forest classification for business bankruptcy prediction pan-european motor vehicles trade sector gives reasonable results. This method gives also a possibility to see how the model reacts to removal of variables by showing the mean decrease of out of the box classification accuracy if a certain variable is dropped and the used set of variables are left intact. The ten most important variables (period, group), ordered by the mean decrease accuracy are listed here:

1. equity / total assets (t-1, FT) ;
2. total liabilities / total assets (t-1, FT);
3. cash flow / total assets (t-1, EF);
4. profit before tax / total assets (t-2, PROF);
5. ROA (t-1, PROF);
6. total liabilities / total assets (t-2, FT);
7. equity / total assets (t-2, FT);
8. cash flow / total assets (t-2, EF);
9. profit before tax / total assets (t-1, PROF);
10. cash flow / total assets (t-2, PROF).

The most important variables are describing the financial structure(FT), the profitability(PROF), and the efficiency(EF). It is notable that from the first ten variables

that had the greatest impact in the accuracy of the model, five were from the t-2 period, which usually has poorer prediction abilities than the t-1 period variables.

The author finds that this thesis makes a good contribution to the business bankruptcy prediction field through application of the random forest classification method. To extend this study, it would not be reasonable to maximise the models for overall prediction accuracy, but to find a way to evaluate possible sum of opportunity cost and the cost of failure, and to optimise such model to minimize the costs. This however means that the author would have to assess a model for evaluating the possible opportunity costs for not giving credit and possible cost for giving credit for failing companies.

Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, **MÄRT PAAL** (sünnikuupäev: 08.09.1986)

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose “Pankroti prognoosimise mudelite koostamine Euroopa autokaubanduse ettevõtete näitel”,

mille juhendaja on teadur Oliver Lukason,

- 1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi Space-Is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
- 1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace´i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus, 25.05.2016